GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:17:10 ; Search time 16.9825 Seconds
(without alignments)
2751.141 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHINRRVQQPPVTATDSFRT.....IEEGGTAASPSEIPFRPMRS 486

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	155.5	6.4	1122	2	G64887	probable tail fibe
2	150.5	6.2	2055	2	T31110	extracellular matr
3	147.5	6.1	1822	2	S33441	EF protein - Strep
4	147	6.1	1156	2	T34852	probable secreted
5	145.5	6.0	971	2	B90835	probable tail fibe
6	145.5	6.0	973	2	C85693	probable membrane
7	144.5	6.0	2232	2	T34434	hypothetical prote
8	138.5	5.7	2155	2	C97523	hypothetical prote
9	138.5	5.7	2155	2	AD2742	conserved hypothet
10	136	5.6	597	2	T35746	hypothetical prote
11	136	5.6	4776	2	E95206	cell wall surface
12	135.5	5.6	1731	2	B98241	hypothetical prote
13	135.5	5.6	1731	2	AB3045	ice nucleation pro
14	135.5	5.6	2479	2	F87386	conserved hypothet
15	135	5.6	718	2	A81122	hypothetical prote
16	135	5.6	1128	1	T08322	plasmid replicatio
17	134	5.6	2271	2	F90073	hypothetical prote
18	133	5.5	1208	2	C82779	hemolysin-type cal
19	132	5.5	1763	2	T17465	rifamycin polyketi
20	132	5.5	1787	2	AG1360	probable tape-meas
21	131.5	5.5	3591	1	S21010	filamentous hemagg
22	131	5.4	556	1	A28185	formate-tetrahydro
23	130	5.4	810	2	B84185	cytochrome-like pr
24	130	5.4	989	2	AE3045	ice nucleation pro
25	130	5.4	1009	2	G98240	hypothetical prote
26	129.5	5.4	703	2	H87360	hypothetical prote
27	129.5	5.4	1541	2	T02831	AAA protein L4171.
28	129	5.3	732	2	E85518	hypothetical prote
29	129	5.3	732	2	B90668	hypothetical prote
30	129	5.3	6713	2	B89921	hypothetical prote
31	128.5	5.3	1026	2	A48995	paracrystalline su
32	128.5	5.3	1073	2	C87374	S-layer protein Rs
33	128.5	5.3	7576	2	T17428	FK506 polyketide s
34	127	5.3	1238	2	T03465	probable exonuclea
35	127	5.3	1326	2	AC3372	kinesin-like prote
36	126.5	5.2	397	2	I39579	nccB protein - Alc
37	126.5	5.2	888	2	D87433	conserved hypothet
38	126.5	5.2	1109	2	A56143	surface-array prot
39	126.5	5.2	2535	2	AC0304	probable hemolysin
40	126	5.2	654	2	C87587	hypothetical prote
41	126	5.2	1147	2	T35781	hypothetical prote
42	125.5	5.2	493	2	S07375	flagellin H-1 - Sa
43	125.5	5.2	653	2	T34976	probable integral
44	125.5	5.2	1065	2	S19482	hypothetical prote
45	125.5	5.2	1430	2	AF0351	probable autotrans

ALIGNMENTS

RESULT 1
G64887
probable tail fiber protein GP37 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: G64887; T09189
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G64887
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1122 <BLAT>
A;Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74454.1; PID:g1787636,
A;Experimental source: strain K-12, substrain MG1655
R;Aiba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai, H
.; Motomura, K.; Nakade, S.; Nakamura, Y.; Nashimoto, H.; Nishio, Y.; Oshima, T.; Saito,
moto, Y.; Horiuchi, T.
DNA Res. 3, 363-377, 1996
A;Title: A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the
A;Reference number: Z16603; MUID:97251357; PMID:9097039
A;Accession: T09189
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 3-1122 <AIB>
A;Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74454.1; PID:g1787636,

Query Match 6.4%; Score 155.5; DB 2; Length 1122;
Best Local Similarity 23.7%; Pred. No. 0.097;
Matches 121; Conservative 60; Mismatches 220; Indels 109; Gaps 21;

```
Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQI 78
           | ||  :  ::: :  :||        | :  |   ||     | :: ||||    :|
Db     117 RNASAVAQNTAAAKKSASDASTSAREAATHAAD---AADSARAASTSAGQAASSAQSASS 173

Qy      79 TALIETRASRLHFEGETPATIAD-TFAKAEKLDRLATTTSGALRATPFAMASLLQYMQPA 137
           :|              | || |  : | ||     | |::|| : :    :: ||    :
Db     174 SA----------GTASTKATEASKSAAAAESSKSAAATSAGAAKTSETNASASLQSAATS 223

Qy     138 INKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAA--S 195
           :              |   | | : | |   :|  ::: : :  |::      | ||  |
Db     224 ASTA----------TTKASEAATSARDAAASKEAAKSS-ETNASSSASSAASSATAAGNS 272

Qy     196 VK---------RHSPSLARQVLDTGVAVQT--YSARNAVRTVLAPALASRPAVQGAVDLG 244
           |          | | : | |        :|   |: :|  |     | ||  |   : :
Db     273 AKAAKTSETNARSSETAAGQSASAAAGSKTAAASSASAASTSAGQASASATAAGKSAESA 332

Qy     245 VSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSA 304
           | |      ||      | :|          |  |  |: || :   |: |   ::
Db     333 ASSASTATTKAGEATEQASAAARS-------ASAAKTSETNAKASETS--AESSKTAAAS 383
```

```
Qy    305 SYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQEMATK 364
          | | || :|    |    | ||    | :|  ||:: :       ||
Db    384 SASSAASSASS--ASASKDEATRQASAAKS--SATTASTKATEAAGS------------ 426

Qy    365 NITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESF-----IQD--TVK--- 414
            || || |: |  | |||  |         : | |:||      ::|  | |
Db    427 -----ATAAAQSKST--AESAATRA---------ETAAKRAEDIASAVALEDASTTKKGI 470

Qy    415 ---STASSTTGYVADQTVKLAKTV---------KDMGGEAITHTGASLRNTVNNLR---- 458
           |:|:::|      | |  |:          ||  |  |   |  | | :| :
Db    471 VQLSSATNSTSETLAATPKAVKSAYDNAEKRLQKDQNGADIPDKGCFL-NNINAVSKTDF 529

Qy    459 --QRPAREADIEEGGTAASPSEIPFRPMRS 486
            :|  |   :     | |    |   |||
Db    530 ADKRGMRYVRVNAPAGATSGKYYPVVVMRS 559
```

RESULT 2
T31110
extracellular matrix binding protein - Abiotrophia defectiva (fragment)
C;Species: Abiotrophia defectiva
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C;Accession: T31110
R;Manganelli, R.; van de Rijn, I.
Infect. Immun. 67, 50-56, 1999
A;Title: Cloning and characterization of emb, a gene encoding the major adhesin of Strep
A;Reference number: Z20988; MUID:99081722; PMID:9864195
A;Accession: T31110
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2055 <MAN>
A;Cross-references: EMBL:AF067776; NID:g3249002; PID:g3249003; PIDN:AAD03320.1
C;Genetics:
A;Gene: emb

  Query Match 6.2%; Score 150.5; DB 2; Length 2055;
  Best Local Similarity 22.2%; Pred. No. 0.44;
  Matches 126; Conservative 73; Mismatches 246; Indels 123; Gaps 21;

```
Qy      3 INRRVQQPPVTA------TDSFRTASDASLA----SSSVRSVSSDQQREINAIADYLTDH 52
          ||   |:| :|        |  ||| ||::|    ::: ::|:| : : :||: :  |
Db    726 INEISQRPDLTREEKQAFMDQVRTARDAAMAKVASAANNQAVTSARDQGLNAVNNLPT-- 783

Qy     53 VFAAHKLPPA-----DSADGQAAVDVHNAQITALIETRASRLHFEGETPATIA------- 100
          | | | |      :|| :      || :||  :  | |     :| | |
Db    784 --PAAKYPEALGHVRQAADAKRQAIRDNANLTAEEQADALRQVDAAQTAAEAAINQNHTN 841

Qy    101 DTFAKAEKLD-RLATTTSGALRATPFAMASLLQY---MQPAINKGDWLPAPLK-PLTPLI 155
          | |||:    :    :   |: | |  :| |     : |||  : |     |       :
Db    842 ATLAKADSDGVKAINDINPQPRSKPAANQALEQVAAAKRQAINNNNQLTDEEKAQAIQQV 901

Qy    156 SGALSGAMDQV-------GTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVL 208
          ||: |  ||       |     : |     : :|     | | :    :    |  |
Db    902 DQALANAKTQVQAANDNNGVNQA-KTAGTTAINNINPQGTQKAQAIAAIEAAEQAKRLEL 960

Qy    209 DTGVAVQTYSARNAVRTVLAPALASRPAVQ---------GAVDLGVSMAGGL-------- 251
            : |     ||:  : | | |:: ||          || | ||:   |:
Db    961 QGRNDLTTEERNNALADLTAKAQAAKDAVNQARNNTGVAGAKDNGVAQIQGINPTAVVKP 1020

Qy    252 -AANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWL-EAYKAIKSASYSGA 309
          | ||      :   :|| :        | |:|  | : :  |   :|  ||  |  :|
Db   1021 DARNA------IDQAARDKEAEFQANTKLTDEEKAAAIKKVQDAARDAKAAIDRAGSNGD 1074

Qy    310 ALNA--------------------GKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALA 349
          ||                      |  |   :  | ||  |  :: : ::|||   |
Db   1075 VNNAVNQGKAAIQAIKALDDSQPSAKDTAKAAIQNAADAKKA--AITANNALTQEEKA-- 1130

Qy    350 GGFAGVGKLQEMATKNITDPATKAAVS---------------QLTNLAGSAAVFAGWTT 393
                  | | : | | ||  :                   |  ::   ||
Db   1131 -----------AAIKQVEDEAAKAQAAVDASRSKADVDRAKDQGLQKISDVPAVQPPKLN 1179

Qy    394 AALTTDPAVKKAESFIQDTVKST---ASSTTGYVADQTVKLAKTVKDMGGEA-ITHTGAS 449
```

```
          |   | |    :: | :    |     :    | ::  |  : : |    | :    |
Db   1180 AIAAVDQAATDKKAVINNDTTLTQEEKEAAIRKVDEEAAKARQAINDATSNADVAAKQAQ 1239

Qy    450 LRNTVNNLRQRP----AREADIEEGGTA 473
            :||: | |    | :| :|:   |
Db   1240 GTQAINNVPQTPAAKNAAKAAVEQAADA 1267
```

RESULT 3
S33441
EF protein - Streptococcus suis
C;Species: Streptococcus suis
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C;Accession: S33441
R;Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.
submitted to the EMBL Data Library, May 1993
A;Description: Repeats in an extracellular protein of wek-pathogenic strains are absent
A;Reference number: S33441
A;Accession: S33441
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1822 <SMI>
A;Cross-references: EMBL:X71880; NID:g298031; PIDN:CAA50714.1; PID:g298032

  Query Match 6.1%; Score 147.5; DB 2; Length 1822;
  Best Local Similarity 20.9%; Pred. No. 0.57;
  Matches 106; Conservative 78; Mismatches 220; Indels 103; Gaps 19;

```
Qy     14 ATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDV 73
          | |:   | |  ||  ::|:   :  : :  :|:   |   |     |    | || :
Db   1222 ARDAVELAKDKELAKEAIRTEEEEATKIVEKLAE---DTRKAIEDNPNLSDEDKQAEIKK 1278

Qy     74 HN---AQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTS----------GAL 120
             |:  | |   | :   | |    :||   ||::  ::|   :          | |
Db   1279 LTDAVAKTLATIRDNADKRTQEAEKAQALAD-LEKAKETQKIADKAAIDRLTILVKDGEL 1337

Qy    121 RATPFAMASLLQYMQPAINK---------------GDWLPAPLKPLTPLISGALSGAMDQ 165
            ||           | | ||                :   |  |  |  :   :: | |
Db   1338 EAT----------KQDAKNKIAKDAAAAKEAIASNPNLTDAEKKTFTDAVDAEVAKANDA 1387

Qy    166 VGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRT 225
          :        ||        |   :  ||  |::        |:|      :|     :|: |: :
Db   1388 ISA-----ATS-----PADVQKEEDAGVAAIAEDVLDAAKQDAKNKIAKDAAAAKEAIGS 1437

Qy    226 VLAPALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGAL---VLGLKDK 282
            | :   |||  |: |    : |          ||  :     |:   ||    :
Db   1438 NPNLTDAEKKTFTDAVDAEVAKANDAISAA---TSPADVQKEEDAGVAAIAEDVLDAAKQ 1494

Qy    283 EPKAQLSEENDWLEAYKAIKSA-----SYSGAALNAGKRMAGLPLDMATDAMGAVRSLVS 337
          : | ::::|:|       | |||     : : |    : |:        ||||: |  | |
Db   1495 DAKNKIAKESD------AAKSAIDANPNLTDAEKESAKKAVDADAKAATDAIDASTSPVE 1548

Qy    338 ASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALT 397
          | |              ||| :     ::: | | : | :::     :|         |:
Db   1549 AQSAEDK---------GVGSI----AQDVLDAAKQDAKNKIAKEVAAA-------KEAID 1588

Qy    398 TDPAVKKAE-SFIQDTVKSTASSTTGYVADQTVKL-AKTVKDMGGEAITHTGASLRNTV- 454
            :| :  ||     :  | : | :||  :    |  : |:: :| |          |:|  |
Db   1589 ANPNLSDAEKEASKKAVDADAKATTDAIDASTSPVEAQSAEDKG-------VGSIRQDVL 1641

Qy    455 ----NNLRQRPAREADIEEGGTAASPS 477
              : : : |:|:|  :    |:|:
Db   1642 DAAKQDAKNKIAKESDAAKSAIDANPN 1668
```

RESULT 4
T34852
probable secreted protein - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T34852
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1999

```
A;Reference number: Z21559
A;Accession: T34852
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1156 <OLI>
A;Cross-references: EMBL:AL035478; PIDN:CAB36606.1; GSPDB:GN00070; SCOEDB:SC2G5.19
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC2G5.19

  Query Match             6.1%;  Score 147;  DB 2;  Length 1156;
  Best Local Similarity   23.1%;  Pred. No. 0.34;
  Matches  122;  Conservative   64;  Mismatches  231;  Indels  112;  Gaps   21;

Qy      13 TATDSFRTA------SDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPAD--- 63
           || |: ||        |   :|::|  :: : ::    |:| :| | |   :|
Db      96 TAEDNRRTVLGAMALSGRGMATASAAALEAGEE----AVAAFL-DGGFKTAELQDLRFNT 150

Qy      64 -SADGQAAVDVHNAQITALIETRASRLH-FEGETPATIADTFAKAEKLDRLATTTSGALR 121
            |    |  : ||  ||| :  :  |  |  :| |  :| |  :|  |     ||
Db     151 LSIQNSAGRGLKNAAQTALSDGTSDALSVFLLDTQFTARNTDERVEVFTILAN------- 203

Qy     122 ATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSG--AMDQVGTKMMDRATGDLH 179
           |   : :| | |:::|         ||    |:    |: |   :  |  |  :
Db     204 ----ASPEVAKYAQRALDEG----------TP---SAIHWFLAIGQYIARARDEETATVD 246

Qy     180 YLSASPDR--------LHDAMAASVK-RHSPSLARQVLDTG----------VAVQTYSAR 220
           | |  :|           |:||| | : : : |::   |           ||    :||
Db     247 QLVAIVEREGKRAQITSDRAVAASDKAKEAAAKAKEAALTAAAEAEAAREDVAKSAAAAR 306

Qy     221 NAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGAL----- 275
           |     | |:| |||  :    | |   | : |     :   :    |  ||
Db     307 KAANAAKGAASAARTAVQAS-----SAAHNAARRSAFAATAAAQAAATAGRAAALAYSAA 361

Qy     276 VLGLKD--KEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVR 333
           |   :|  |   |:|: |     | || |:|    ||  | :  |   :  |  |  :
Db     362 VAAARDASKTKAARLAAEGARNAAAKARKAAQALAAAQTATQAAAAAGISAAATARDSAA 421

Qy     334 S------LVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAA--VSQLTNLAGSA 385
           :          ||   |:  |:|  |      |           |||||     | | | ||
Db     422 AAQQAAVAAQASGAAQSEAAVARAAAAEADAQAAR-------ATKAANRAQSLANTAASA 474

Qy     386 AVFA------------------GWTTAALTTDPAVKKAESFIQDTVKSTASSTTGYVAD 426
           | |                    ||   |    ||:::  |:|   |:      |
Db     475 AAAARKAADSAAAHAEKAADAADAAADAAGEADDYANKAKAWAADSV--AAAELAAKAVD 532

Qy     427 QTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAAS 475
           :    ::   | : |       ::   |:  | ||:  |    |:
Db     533 DARAVEAAAREAEAEKLAH---DTEQSLAEAREMAAAEAEDREAARNAA 578


RESULT 5
B90835
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: B90835
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B90835
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35073.1; PID:g13361114; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1650

  Query Match             6.0%;  Score 145.5;  DB 2;  Length 971;
  Best Local Similarity   22.5%;  Pred. No. 0.33;
  Matches  107;  Conservative   52;  Mismatches  232;  Indels   85;  Gaps   15;

Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQI 78
           | ||  :  ::: :  :||        | : ||   ||     | :: ||||    :|
Db     115 RNASAVAQNTAAAKKSASDASTSAREAATHATD---AADSARAASTSAGQAASSAQSASS 171

Qy      79 TALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQPAI 138
           :|          | || |   | | :  : |  ||     |    |:: |  | |
Db     172 SA----------GTASTKATEASKSAAAAESSKSAAATSAGAAKTSETNAAVSQ--QSAA 219

Qy     139 NKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAA--SV 196
                   |   | | | | |   :|   ::: :    |::       | ||  |
Db     220 TSA-------STATTKASEAASSARDASASKEAAKSS-ETSAASSASSAASSATAAGNSA 271

Qy     197 K---------RHSPSLARQVLDTGVAVQTYSA--RNAVRTVLAPALASRPAVQGAVDLGV 245
           |         : | : | |        :| :|   :|  |    | ||  |   : :
Db     272 KAAKTSETNAKSSETAAEQSASAAAGSKTAAALSASAASTSAGQASASATAAGKSAESAA 331

Qy     246 SMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSAS 305
           | |       ||      |  :                 |  |  |: |:     |  |||
Db     332 SSASTATTKAGEATEQASAAASSASAAKTSETNAKASETSAESSKTAAASSASSAASSAS 391

Qy     306 YSGAA-------LNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKL 358
           : |:         :|  |  |       ||:| |:  :   : |  ::     |
Db     392 SASASKDEATRQASAAKSSATTASTKATEAAGSATAAAQSKSTAESAATRA--------- 442

Qy     359 QEMATKNITDPA----------TKAAVSQLTNLAGSAAVFAGWTTAALTTDP-AVKKAES 407
           | | |   | |           ||  : ||::   |        |: :|    | ||| |
Db     443 -ETAAKRAEDIASAVALEDASTTKKGIVQLSSATNS-------TSESLAATPKAVKAAYE 494

Qy     408 F------IQDTVK--------STASSTTGYVADQTVKLAKTVKDMGGEAITHTGAS 449
                   ||           | |:::|  :   | |  |    |:       |:
Db     495 LANGKYTAQDATTAQKGIVQLSNATNSTSEMLAATPKSVKAAYDLANGKYTAQDAT 550


RESULT 6
C85693
probable membrane protein of prophage CP-933X Z1918 [imported] - Escherichia coli (strai
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: C85693
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85693
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-973 <STO>
A;Cross-references: GB:AE005174; NID:g12514847; PIDN:AAG56007.1; GSPDB:GN00145; UWGP:Z15
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1918

  Query Match             6.0%;  Score 145.5;  DB 2;  Length 973;
  Best Local Similarity   22.5%;  Pred. No. 0.33;
  Matches  107;  Conservative   52;  Mismatches  232;  Indels   85;  Gaps   15;

Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQI 78
           | ||  :  ::: :  :||        | : ||   ||     | :: ||||    :|
Db     117 RNASAVAQNTAAAKKSASDASTSAREAATHATD---AADSARAASTSAGQAASSAQSASS 173

Qy      79 TALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQPAI 138
           :|          | || |   | | :  : |  ||     |    |:: |  | |
Db     174 SA----------GTASTKATEASKSAAAAESSKSAAATSAGAAKTSETNAAVSQ--QSAA 221

Qy     139 NKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAA--SV 196
                   |   | | | | |   :|   ::: :    |::       | ||  |
Db     222 TSA-------STATTKASEAASSARDASASKEAAKSS-ETSAASSASSAASSATAAGNSA 273

Qy     197 K---------RHSPSLARQVLDTGVAVQTYSA--RNAVRTVLAPALASRPAVQGAVDLGV 245
           |         : | : | |        :| :|   :|  |    | ||  |   : :
```

```
Db    274 KAAKTSETNAKSSETAAEQSASAAAGSKTAAALSASAASTSAGQASASATAAGKSAESAA 333

Qy    246 SMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSAS 305

Db    334 SSASTATTKAGEATEQASAAASSASAAKTSETNAKASETSAESSKTAAASSASSAASSAS 393

Qy    306 YSGAA-------LNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKL 358

Db    394 SASASKDEATRQASAAKSSATTASTKATEAAGSATAAAQSKSTAESAATRA--------- 444

Qy    359 QEMATKNITDPA----------TKAAVSQLTNLAGSAAVFAGWTTAALTTDP-AVKKAES 407

Db    445 -ETAAKRAEDIASAVALEDASTTKKGIVQLSSATNS-------TSESLAATPKAVKAAYE 496

Qy    408 F------IQDTVK--------STASSTTGYVADQTVKLAKTVKDMGGEAITHTGAS 449

Db    497 LANGKYTAQDATTAQKGIVQLSNATNSTSEMLAATPKSVKAAYDLANGKYTAQDAT 552
```

RESULT 7
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T34434
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2232 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/

Query Match 6.0%; Score 144.5; DB 2; Length 2232;
Best Local Similarity 18.7%; Pred. No. 1.2;
Matches 99; Conservative 81; Mismatches 261; Indels 89; Gaps 11;

```
Qy      3 INRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPP- 61

Db   1306 IGSTVTQPS-TVSGSNSSGSTVTIGSSEASTSGSSFKTSPSSISPVPT-----SSPIPST 1359

Qy     62 --ADSADGQAAVDVHNAQITALI----------------------ETRASRLHFEGETP 96

Db   1360 TFASSTSGSTISDVSSVSTTSLAPLSSSLPSTVPSSTQSFSSTSEGSSKASSSPVPSQTS 1419

Qy     97 ATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQPAINK--GDWLPAPLKPLTPL 154

Db   1420 STPTNPTGSTESSTLLSSTISGSTQHTTMSKAS-SGSTSPSTNSQTGSTVTMGSSSTSGV 1478

Qy    155 ISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAA-------------SVKRHSP 201

Db   1479 STSSASSTQPQMSTSQGSSAGSTVASSTASPAASSTAPSSTGTMSSTSSGTVGSTISESS 1538

Qy    202 SLARQVLDTGVAV-------------------------QTYSARNAVRTVLAPALASRPA 236

Db   1539 TTASASSQTGSTVTMGSSSTSGVSTSSASSTQPQMSTSQGSSAGSTVASSTAGLVSTSTV 1598

Qy    237 VQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLE 296

Db   1599 PSSTGTMGSTSSGTVGSTISESSTTASASS---QTGSTVTMGSSSTSGVSTSSASSTQPQ 1655

Qy    297 AYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVG 356

Db   1656 MSTSQGSSAGSTVASSTTGLVSTSTVPSSTGTMGSTSSGTVGSTISESSTA-ASASSQTG 1714

Qy    357 KLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKST 416

Db   1715 STVTMGSSSTSGVSTSSASSGQPQMSTSQGSSAGSTVVSSTASPAASSTAPSSTGTMSST 1774

Qy    417 ASSTTGYVADQTVKLAKTVKDMG----------GEAITHTGASLRNTV 454

Db   1775 SSGTVGSTMSQSSTAASTTSHTGSTVTLGSSSTSSNQMSTSQGSSVGSTV 1824
```

RESULT 8
C97523
hypothetical protein AGR_C_2490 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: C97523
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: C97523
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2155 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87140.1; PID:g15156408; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_2490
A;Map position: circular chromosome

Query Match 5.7%; Score 138.5; DB 2; Length 2155;
Best Local Similarity 20.0%; Pred. No. 2.6;
Matches 119; Conservative 82; Mismatches 194; Indels 199; Gaps 23;

```
Qy     17 SFRTASDASLASSSVRSVSSDQQREINAIADYL-------------------------TD 51

Db    218 ALRLAEPETVASDRIMSVGQAVRREVSAMNDGIERTIARATELETLVHSEVNALERSYAD 277

Qy     52 HVFAAHKLPPADSADGQAAVDVHNAQITALI---------------ETRASRLHFEGETP 96

Db    278 NELRVRSLVQELTAERDAIVN-HAERIRSSIVGAQEQIKEELSIVGEELSMRIATTGEAF 336

Qy     97 ATIADTFAKA--EKLDRLATTTSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPL 154

Db    337 ASMIDTRSAALLEK-SRASTEAMGSL------IAAKTENLLQALNS--------------- 375

Qy    155 ISGA------------LSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPS 202

Db    376 -SGSTISNEFDMRLHNLTSTLDERGEVLLERFAIHASTLDSGVESLNSALEERTRQLNET 434

Qy    203 LARQVLDTGVAVQTYSARNAVR-----------------TVLAPALASRPAVQGAVD-- 242

Db    435 LSARSLELN--------RNIERGQQVIGGSLDTVLDKLSTTLEEKGLSFRQSLQSTADDA 486

Qy    243 -LGVSMAGGL----------AANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEE 291

Db    487 IMDLDLRSGLYEERMQATVGQVNSAFDEHVAQFASAFDQRAGSL-----DSKLMESLARI 541

Qy    292 NDWL----EAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLA 347

Db    542 NETVAGGSEALDTI---------LTSGLERIG------------------STMTDQSLA 573

Qy    348 LAGGFAGVGKLQEM-----------------ATKNITDPATKAAVSQLTNLA-GSAAVF 388

Db    574 LA---TALGTGQEMLENALESRTQAFSDAIGQRTAEITDAFTNSHAKIDTVLAERSNALF 630

Qy    389 AGWTTAALTTDPAVKKAESFIQDTVKST----------ASSTTGYVADQTVKLAKTVKD 437

Db    631 GALSASQDRFDEALASRSLAITGSVSGTAEHLAAMLDERAAAINSVVADVERRLTETLET 690

Qy    438 MGGEAITHTGASLRNTVNN-LRQRPA----------READIEEGGTAASPSEI 479

Db    691 RAA-AITGAVSGIEDRISDTLESRTAALHDVVSGAESRIADTLDGRTAALSSAI 743
```

```
RESULT 9
AD2742
conserved hypothetical protein Atu1348 [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AD2742
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AD2742
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2155 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL42354.1; PID:g17739760; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu1348
A;Map position: circular chromosome

  Query Match             5.7%;  Score 138.5;  DB 2;  Length 2155;
  Best Local Similarity   20.0%;  Pred. No. 2.6;
  Matches  119;  Conservative   82;  Mismatches  194;  Indels  199;  Gaps   23;

Qy      17 SFRTASDASLASSSVRSVSSDQQREINAIADYL------------------------TD 51
           :  |  |    ::||   :  ||     :||::|: | :                        |
Db     218 ALRLAEPETVASDRIMSVGQAVRREVSAMNDGIERTIARATELETLVHSEVNALERSYAD 277

Qy      52 HVFAAHKLPPADSADGQAAVDVHNAQITALI--------------ETRASRLHFEGETP 96
           :       |     :|:   |  |: |   :| : |                |  : |:    ||
Db     278 NELRVRSLVQELTAERDAIVN-HAERIRSSIVGAQEQIKEELSIVGEELSMRIATTGEAF 336

Qy      97 ATIADTFAKA--EKLDRLATTTSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPL 154
           |::  ||  :  |   ||   | :|    |:|       :|:  : :   |:|
Db     337 ASMIDTRSAALLEK-SRASTEAMGSL------IAAKTENLLQALNS-------------- 375

Qy     155 ISGA------------LSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPS 202
           ||:              |:  :|: |  :::|       | :  : |: |:    :: : :
Db     376 -SGSTISNEFDMRLHNLTSTLDERGEVLLERFAIHASTLDSGVESLNSALEERTRQLNET 434

Qy     203 LARQVLDTGVAVQTYSARNAVR-----------------TVLAPALASRPAVQGAVD-- 242
           |: : |:              ||  |               |:     |: | ::|    |
Db     435 LSARSLELN--------RNIERGQQVIGGSLDTVLDKLSTTLEEKGLSFRQSLQSTADDA 486

Qy     243 -LGVSMAGGL----------AANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEE 291
            : : :  ||           |: |   :    |   || |:|      | :    |:
Db     487 IMDLDLRSGLYEERMQATVGQVNSAFDEHVAQFASAFDQRAGSL-----DSKLMESLARI 541

Qy     292 NDWL----EAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLA 347
           |: :    ||   |          | :|    |                |::|   ||
Db     542 NETVAGGSEALDTI---------LTSGLERIG-----------------STMTDQSLA 573

Qy     348 LAGGFAGVGKLQEM-----------------ATKNITDPATKAAVSQLTNLA-GSAAVF 388
           ||     :|  |||                  |  |||  | :     | ||  | |:|
Db     574 LA---TALGTGQEMLENALESRTQAFSDAIGQRTAEITDAFTNSHAKIDTVLAERSNALF 630

Qy     389 AGWTTAALTTDPAVKKAESFIQDTVKST----------ASSTTGYVADQTVKLAKTVKD 437
           : :    | |:      |  :|  |           |::    |||   :| :|::
Db     631 GALSASQDRFDEALASRSLAITGSVSGTAEHLAAMLDERAAAINSVVADVERRLTETLET 690

Qy     438 MGGEAITHTGASLRNTVNN-LRQRPA----------READIEEGGTAASPSEI 479
           |||   : : : ::: |  | |           | ||  :| |||  | |
Db     691 RAA-AITGAVSGIEDRISDTLESRTAALHDVVSGAESRIADTLDGRTAALSSAI 743


RESULT 10
T35746
hypothetical protein SC7H2.15c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T35746
R;Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M
submitted to the EMBL Data Library, August 1999
A;Reference number: Z21588
A;Accession: T35746
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-597 <SAU>
A;Cross-references: EMBL:AL109732; PIDN:CAB52057.1; GSPDB:GN00070; SCOEDB:SC7H2.15c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC7H2.15c

  Query Match             5.6%;  Score 136;  DB 2;  Length 597;
  Best Local Similarity   21.5%;  Pred. No. 0.66;
  Matches  116;  Conservative   75;  Mismatches  226;  Indels  122;  Gaps   21;

Qy       2 HINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIAD--YLTDHVFAAHKL 59
           | :||:::  ||  |           :|: :| ::|:  |  ::  |  :    |
Db       4 HFDRRLRKAAVTTT----------VAAVAVAALSASQAPDVTADGNGRQTTADNAPTSDT 53

Qy      60 PPADSADGQAAV-----DVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLAT 114
           || :|| | :         :::         | |||   |   |||: | : |||   | |
Db      54 PPEESATGNSRYYTDLPPLNSPSPAPTTGTPASRGASEAGIPATVLDAYKKAESELRRAK 113

Qy     115 TTSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGAL-----------SGA 162
                 |: : : :  ::    :|  : |    :  :|   |             :|
Db     114 PGCN----LPWQLLAAIGKVESGQARGGRVDANGTTIGRIIGPQLDGNGFALIKDTDNGV 169

Qy     163 MDQVGTKMMDRATGDLHYLSAS--------------PDRLHDAMAAS---VKRHSPSLA 204
           |  |    | | | : :: ::              |: ::||  |:    : |:|  ||
Db     170 YD--GNSSYDNAVGPMQFIPSTWAWAGRDGNSDGKEDPNNVYDAALAAGHYLCRNSWDLA 227

Qy     205 RQVLDTGVAVQTY-SARNAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFGNRL-- 261
           |  |    |: :| :::: : |||:      |       |    |:  | : :|  | |
Db     228 DQA-DLKRAILSYNNSQDYLHTVLSWLEYYRKGTHEIPDGTGSLPVGRSDDATPGTRSPG 286

Qy     262 --LSVQSRDHQRGGALVL-GLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMA 318
               ||    | |    :| |                                ||    : |
Db     287 RGTGASSRPQSPGTATPSPGRSPSKPPA--------------------SGRPTDPGSPST 326

Qy     319 GLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVS-- 376
             |     ||       :     ::     | | ||: :       | :  |    |||  :
Db     327 S-PSTPPTDPAPPTETPTPTDTVHHLENAGATGFSAIAG-DTFAERISTRAETKAGEAVG 384

Qy     377 ----QLTNLAGSAAVFAGW-TTAALTTD-------PAVKKAESFIQDTVKSTASSTT--- 421
               : | |  : | | |    ||:||:        ||:   |      ||::|    |
Db     385 KVRVRFTVLGDTDATFTGGEKVAAVTTNASGVAVAPALVAGEQTGGFTVRATVIGRTVPG 444

Qy     422 -GYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEI 479
            | |  | ::|         :|:| || :           |     ||  | | |:
Db     445 LDYKATVTERVA--------DALTSTGET-------------EPTCTPGGEFADPIEV 481


RESULT 11
E95206
cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain TI
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: E95206
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: E95206
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-4776 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK75846.1; PID:g14973269; GSPDB:GN00164; TIGR:SP4
```

```
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1772

  Query Match             5.6%;  Score 136;  DB 2;  Length 4776;
  Best Local Similarity  20.0%;  Pred. No. 11;
  Matches   94;  Conservative   68;  Mismatches  272;  Indels   36;  Gaps    9;

Qy      13 TATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVD 72
           :|::|  |:: || ::|:  | |:      :  |        |:     : ||   |:
Db    4306 SASESASTSASASASTSASASASTSASASASTSAS-------ASASTSASVSASTSASES 4358

Qy      73 VH-NAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLL 131
            :|  :|     ||       |: :| |   |     :  :|: | :   :  | ||
Db    4359 ASTSASASASTSASASASTSASESASTSASASASTSASESASTSASASASTSASASAS-- 4416

Qy     132 QYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDA 191
                        |      |  : | :  |      |:     |||      |
Db    4417 ------------TSASASASTSASASASTSASASASTSASASAS-----TSASASASTSA 4459

Qy     192 MAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGAVDLGVSMAGGL 251
            |::    | | :     |  :    :: :|  :  | | ||  |   :  :  | :
Db    4460 SASASTSASASAS-----TSASASASTSASASASTSASASASTSA-SASASISASESAST 4513

Qy     252 AANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAAL 311
           :|:|   :   |  :        |           |  |      |:     ||| | :|
Db    4514 SASAS-ASTSASASASTSASVSASTSASASASTSASASASISASESASTSASASASTSAS 4572

Qy     312 NAGKRMAGLPLDMATDAMGAVRSLVSAS-SLTQNGLALAGGFAGVGKLQEMATKNITDPA 370
            :    |      :  |  :: :  |||  | : :    |   |          :|     :
Db    4573 ASASTSASASASTSASASASISASESASTSASASASTSASASASTSASASASTSASASAS 4632

Qy     371 TKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASSTTGYVADQTVK 430
           | |:  |  |: : ||:  |   :|: :   :   :  |     |  ::||:   :  |
Db    4633 TSASASASTSASASASTSAS-ASASTSASASASTSASASASTSASASASTSASASASTSA 4691

Qy     431 LAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEIP 480
           |          |  |    ||   :|:|       :      |  |  |  :|
Db    4692 SASASTSASASASTSASASASTSVSNSANHSNSQVGNTSGSTGKSQKELP 4741


RESULT 12
B98241
hypothetical protein AGR_L_1764 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: B98241
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: B98241
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1731 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89452.1; PID:g15159316; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1764
A;Map position: linear chromosome

  Query Match             5.6%;  Score 135.5;  DB 2;  Length 1731;
  Best Local Similarity  24.4%;  Pred. No. 2.9;
  Matches  125;  Conservative   69;  Mismatches  217;  Indels  101;  Gaps   26;

Qy       9 QPPVTATD---SFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSA 65
           |  | :||   :| |   |:| ||::  :| |       :|   |  : |      :  :
Db     907 QLAVLSTDDIATFSTRDMAALGSSAIAGLSKD------TVASLTTAQIGALSMAGISGLS 960

Qy      66 DGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPF 125
             || |  :   |:  |  |: : |  :      ::|  | :       | : :||   |
Db     961 TGQIAA-LTGDQLNVLTNTQIAALTSKQVAAFDVSDITALSTG-QIAALSAAGAAGLTTD 1018

Qy     126 AMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSA-- 183
           :|:|      |:  |        | :  |          | :  |     :|||:  ||:
Db    1019 QIAALSTDQVGAMTSGQIAALSAKQIAAL-------GTDDIAT----FSTGDIAALSSNA 1067

Qy     184 ----SPDRLHDAMAASVKRHSPSLARQVLDTG-VAVQTYSARNAVRTVLAPALASRP-AV 237
               | | :     | :   | |      | || :|  |     |  :      || |:  |
Db    1068 VAGLSRDTVASLTTAQIAALS-SAGISGLGTGQIAGLTSEQVNVLTNAQISALTSKQVAA 1126

Qy     238 QGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEA 297
           |:    |  :||                  | | | ||   :  | ||     :
Db    1127 LEVTDIASLSAAQIAA----------------IGAAGVAGL-TTDQIAALS-----ISQ 1163

Qy     298 YKAIKSASYSGAALNAGKRMAGLPLD---------MATDAMGAVRSLVSA--SSLTQNGL 346
           :|: ||    ||||: |::| |  |          ||    ||:  | ::  :|||   :
Db    1164 VEALTSAQI--AALNS-KQIAALSADDLAIFTTAEMAAIGSGAISGLPASTIASLTTAQI 1220

Qy     347 ALAGGFAGVG----KLQEMATKNITDPATKAAVSQLT-------NLAG----SAAVFAGW 391
           |  |  |  |    ::  :  |  :  |  | | :  ||       :::|    |:|  |
Db    1221 AALGAAAVSGLTTDQIAALGTGQV-DALTNAQIGALTSKQVTALSVSGIGSLSSAQIAAL 1279

Qy     392 TT---AALTTDPAVKKAESFIQ--DTVKSTASSTTGYVA----DQTVKLAKTVKDMGGEA 442
           :|   | ||||     : | ::   :|:  | |:     |    | :      :  :|  |
Db    1280 STAGVAGLTTDQIAALSTSQVEALTSVQIAALSSKQIAALGADDLDIFTTAEIASIGSSA 1339

Qy     443 ITHTGASLRNTVNNLRQRPAREADIEEGGTAA 474
           :    ||   |: :|       | |   ||||
Db    1340 VAGLSAS---TIASL-----TTAQIAALGTAA 1363


RESULT 13
AB3045
ice nucleation protein homolog [imported] - Agrobacterium tumefaciens (strain C58, Dupon
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AB3045
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AB3045
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1731 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44776.1; PID:g17742414; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: ina
A;Map position: linear chromosome

  Query Match             5.6%;  Score 135.5;  DB 2;  Length 1731;
  Best Local Similarity  24.4%;  Pred. No. 2.9;
  Matches  125;  Conservative   69;  Mismatches  217;  Indels  101;  Gaps   26;

Qy       9 QPPVTATD---SFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSA 65
           |  | :||   :| |   |:| ||::  :| |       :|   |  : |      :  :
Db     907 QLAVLSTDDIATFSTRDMAALGSSAIAGLSKD------TVASLTTAQIGALSMAGISGLS 960

Qy      66 DGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPF 125
             || |  :   |:  |  |: : |  :      ::|  | :       | : :||   |
Db     961 TGQIAA-LTGDQLNVLTNTQIAALTSKQVAAFDVSDITALSTG-QIAALSAAGAAGLTTD 1018

Qy     126 AMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSA-- 183
           :|:|      |:  |        | :  |          | :  |     :|||:  ||:
Db    1019 QIAALSTDQVGAMTSGQIAALSAKQIAAL-------GTDDIAT----FSTGDIAALSSNA 1067

Qy     184 ----SPDRLHDAMAASVKRHSPSLARQVLDTG-VAVQTYSARNAVRTVLAPALASRP-AV 237
               | | :     | :   | |      | || :|  |     |  :      || |:  |
Db    1068 VAGLSRDTVASLTTAQIAALS-SAGISGLGTGQIAGLTSEQVNVLTNAQISALTSKQVAA 1126
```

```
Qy    238 QGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEA 297
Db   1127 LEVTDIASLSAAQIAA-----------------IGAAGVAGL-TTDQIAALS-----ISQ 1163

Qy    298 YKAIKSASYSGAALNAGKRMAGLPLD---------MATDAMGAVRSLVSA--SSLTQNGL 346
Db   1164 VEALTSAQI--AALNS-KQIAALSADDLAIFTTAEMAAIGSGAISGLPASTIASLTTAQI 1220

Qy    347 ALAGGFAGVG----KLQEMATKNITDPATKAAVSQLT-------NLAG----SAAVFAGW 391
Db   1221 AALGAAAVSGLTTDQIAALGTGQV-DALTNAQIGALTSKQVTALSVSGIGSLSSAQIAAL 1279

Qy    392 TT---AALTTDPAVKKAESFIQ--DTVKSTASSTTGYVA----DQTVKLAKTVKDMGGEA 442
Db   1280 STAGVAGLTTDQIAALSTSQVEALTSVQIAALSSKQIAALGADDLDIFTTAEIASIGSSA 1339

Qy    443 ITHTGASLRNTVNNLRQRPAREADIEEGGTAA 474
Db   1340 VAGLSAS---TIASL-----TTAQIAALGTAA 1363


RESULT 14
F87386
conserved hypothetical protein CC1106 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: F87386
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87386
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2479 <STO>
A;Cross-references: GB:AE005673; NID:g13422414; PIDN:AAK23090.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1106

  Query Match             5.6%;  Score 135.5;  DB 2;  Length 2479;
  Best Local Similarity   23.1%;  Pred. No. 4.8;
  Matches  113;  Conservative   67;  Mismatches  211;  Indels   99;  Gaps   19;

Qy     13 TATDSFRTASDASLASSSVRSVSSDQQREINAI-ADYLTDHVFAAHKLPPADSADGQAAV 71
Db   1766 TQIAALQTSQLAALATTQAQTLSTTQLGALNATQVGALTTQAFSALDATQVGALTTTQVK 1825

Qy     72 DVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLL 131
Db   1826 GLTNSQLSALTTTDVGQL----------ADT-----QVGALSATQLGALTATAFSALDAT 1870

Qy    132 QYMQPAINKGDWLPA-PLKPLTPLISGALS-----------GAMDQVGTKMMDR----- 173
Db   1871 QTAALTTTQVKGLTATQLKGLTTTDIGDLADTQVGALSATQLGALTATGFSSLNATQVAV 1930

Qy    174 -----------------ATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQT 216
Db   1931 LSATQIKGLTADQLKGLSTTDIGELSTT--QITALTADQLKNLSATNISAMADTQVAALS 1988

Qy    217 YSARNAVRTVLAPALASRPAVQ-GAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGAL 275
Db   1989 NTQIQALSTTQLKGLSVTDIDELSATQVGALSAAQIAALASTQIQELSTTQ--------- 2039

Qy    276 VLGLKDKEPKAQLSEENDWLEAYK--AIKSASYSGAALNAGKRMAGLPL----DMATDAM 329
Db   2040 IGGLSATQIGAVSATNVSLLNATQVGALTTTQVKGLTV---AQLAGLSTTNIGDLADTQV 2096

Qy    330 GAVRSLVSASSLTQNGLALAGGFA-----GVGKLQEMATKNITDPATKA-AVSQLTNLAG 383
Db   2097 GALSS-------TQLGALTATGFSTLNGTQVAALNSTQLKGLTTTQLKGLTTTDISELAD 2149

Qy    384 S--AAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASSTT---GYVADQTVKLAKTVKDM 438
Db   2150 TQVAALTAAQIAALSATNVSALNA-------TQVTALTTTQVKGLTVDQLKGL--TTTDI 2200

Qy    439 GGEAITHTGA 448
Db   2201 GELANTQIGA 2210


RESULT 15
A81122
hypothetical protein NMB1108 [imported] - Neisseria meningitidis (strain MC58 serogroup
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: A81122
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.,
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: A81122
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-718 <TET>
A;Cross-references: GB:AE002460; GB:AE002098; NID:g7226335; PIDN:AAF41499.1; PID:g72263
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1108

  Query Match             5.6%;  Score 135;  DB 2;  Length 718;
  Best Local Similarity   22.4%;  Pred. No. 0.97;
  Matches  119;  Conservative   76;  Mismatches  224;  Indels  112;  Gaps   25;

Qy     14 ATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDV 73
Db    161 ATEGAQQIKD--LALELVEKNGGTHDKALDLISGMMTTGLNFAQTKNEAQAAYAFALASE 218

Qy     74 HNAQITA-LIET--------RASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATP 124
Db    219 GSGEDTAKLIKTLKDGGMSGKDLQLGLEHVLQSGLDGTFEVRDMVRELPSLLSAAQQAGM 278

Qy    125 FAMAS---LLQYMQPAINKGDWLPAPL---------KPLTPLISGAL---------SGAM 163
Db    279 NGVGGLDYLLSLLQSAANKSG-SPAEAATNVQNLLSKTLSPDTIGRLKKMANPNDPKKGV 337

Qy    164 DQVGTKMMDRATGD--LHYLSASPDRLHDAMAASVKRHSPSLARQVL-DTGVAVQTYSAR 220
Db    338 DWIGSVVQGKQNGENAVQVLS----RLADAMLVKDKQYQDYKKRAAAGDKTAAEQANMLK 393

Qy    221 NAVRTVLAPALASRPAVQGAVDL----------GVSMAGGLAANAGFGNRLLSVQSRDH 269
Db    394 GALLAQLLPDLQAKQGLLAATDMTQIREYMASLAGVTLDNGKIAKNNEA-RMLSAAAQQE 452

Qy    270 QRGGALVLGLKDKEPKAQLSEE-----NDWLEAYKAIKSASYSGAALNAGKRMAGLPLDM 324
Db    453 Q-----------QESLAMLRESLTGTLVDMETSFKKL-AAEYPNATL------------- 487

Qy    325 ATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQEMATKNI-TDPATKAAVSQLTNLAG 383
Db    488 ---ALQALTTAATAASAAM--LLTAGGGKGAGFLKDVGSKALGWGKASAGGVAAGATAAG 542

Qy    384 SAAVFAGWTTAA---LTTDPAVKKAESFIQDTVKSTASSTTGYVADQTVKLAKTVKDMGG 440
Db    543 GKLL--SWGKSAGSGLMNNPALVKRAGLLGMLLYSES------LGDGTLP-----KGLRG 589

Qy    441 EAITHTGASLRNTV--NNLRQRPAREADIEEGGT---AASPSEIPFRPMRS 486
Db    590 ---TKTTPEMINRLKNNGIRFEPAPKREQARGGVPQYLAAPSAQPTDKMLS 637


Search completed: January  2, 2003, 15:21:12
```

Job time : 29.9825 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:19:06 ; Search time 8.50874 Seconds
(without alignments)
1084.696 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINQSAQQPPGVAMESFRT.....EEGGISAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2407	100.0	487	10	US-09-825-414-66	Sequence 66, Appl
2	1869.5	77.7	486	10	US-09-825-414-7	Sequence 7, Appli
3	147	6.1	1463	9	US-09-971-536-69	Sequence 69, Appl
4	146.5	6.1	2122	9	US-09-813-214A-9	Sequence 9, Appli
5	140	5.8	2285	10	US-09-932-183A-2	Sequence 2, Appli
6	138	5.7	2993	9	US-09-738-626-6239	Sequence 6239, Ap
7	132	5.5	1026	10	US-09-379-931-7	Sequence 7, Appli
8	130.5	5.4	2368	10	US-09-815-242-5635	Sequence 5635, Ap
9	130.5	5.4	2368	10	US-09-815-242-12389	Sequence 12389, A
10	129.5	5.4	2076	10	US-09-815-242-5815	Sequence 5815, Ap
11	129.5	5.4	2186	10	US-09-815-242-12913	Sequence 12913, A
12	129	5.4	2344	10	US-09-815-242-12713	Sequence 12713, A
13	127.5	5.3	2478	10	US-09-815-242-5816	Sequence 5816, Ap
14	127.5	5.3	2478	10	US-09-815-242-12967	Sequence 12967, A
15	125.5	5.2	1222	10	US-09-137-531-15	Sequence 15, Appl
16	125.5	5.2	1252	10	US-09-137-531-9	Sequence 9, Appli
17	125	5.2	2086	10	US-09-815-242-5639	Sequence 5639, Ap
18	125	5.2	5795	10	US-09-815-242-12610	Sequence 12610, A
19	124	5.2	628	10	US-09-841-786-5	Sequence 5, Appli
20	124	5.2	3241	10	US-09-841-786-1	Sequence 1, Appli
21	123.5	5.1	2434	10	US-09-815-242-5835	Sequence 5835, Ap
22	123.5	5.1	6281	10	US-09-815-242-12996	Sequence 12996, A
23	120.5	5.0	810	9	US-09-712-363-281	Sequence 281, App
24	119.5	5.0	778	10	US-09-815-242-10728	Sequence 10728, A
25	119.5	5.0	1029	10	US-09-815-242-5885	Sequence 5885, Ap
26	119.5	5.0	1048	10	US-09-815-242-13083	Sequence 13083, A
27	119	4.9	600	10	US-09-287-849-22	Sequence 22, Appl
28	118	4.9	596	9	US-10-063-547-100	Sequence 100, App
29	118	4.9	596	12	US-10-006-867-100	Sequence 100, App
30	118	4.9	596	12	US-10-052-586-310	Sequence 310, App
31	118	4.9	647	10	US-09-841-132-436	Sequence 436, App
32	117.5	4.9	1215	10	US-09-815-242-5908	Sequence 5908, Ap
33	117.5	4.9	1269	10	US-09-815-242-13113	Sequence 13113, A
34	117.5	4.9	3158	10	US-09-815-242-12611	Sequence 12611, A
35	114	4.7	573	10	US-09-828-574-13	Sequence 13, Appl
36	114	4.7	729	10	US-09-287-849-2	Sequence 2, Appli
37	113	4.7	1054	10	US-09-798-042-87	Sequence 87, Appl
38	112	4.7	660	12	US-10-007-693-139	Sequence 139, App
39	112	4.7	1879	9	US-09-971-536-70	Sequence 70, Appl
40	111	4.6	1488	9	US-09-738-626-5495	Sequence 5495, Ap
41	110	4.6	792	9	US-10-025-380-1127	Sequence 1127, Ap
42	109.5	4.5	4999	9	US-09-976-059-14	Sequence 14, Appl
43	108.5	4.5	764	10	US-09-815-242-5143	Sequence 5143, Ap
44	108	4.5	458	9	US-09-738-626-6104	Sequence 6104, Ap
45	108	4.5	596	10	US-09-287-849-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-825-414-66
; Sequence 66, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 66
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. tomato
US-09-825-414-66

Query Match 100.0%; Score 2407; DB 10; Length 487;
Best Local Similarity 100.0%; Pred. No. 2.9e-180;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60

Qy      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120

Qy     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180
```

```
Qy    181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240

Qy    241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300

Qy    301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360

Qy    361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420

Qy    421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480

Qy    481 PFQLRRL 487
          |||||||
Db    481 PFQLRRL 487
```

RESULT 2
US-09-825-414-7
; Sequence 7, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
US-09-825-414-7

Query Match 77.7%; Score 1869.5; DB 10; Length 486;
Best Local Similarity 78.4%; Pred. No. 2.4e-138;
Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;

```
Qy      1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
          ||||:  ||||  | :||||||||||||||||||:   |:: || |||  ||||||:
Db      1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60

Qy     61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
            | | : |   || || ||:||||:||: |||||||||:|||||| |||||| | |
Db     61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120

Qy    121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180
          ||||  ||:|||||||||||||  ||||||||||||||||||||||||||| |||||
Db    121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180

Qy    181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240
          || |||:|||||| |||||||:| |||:| |:||||:|| | ||||||||||||:||||
Db    181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240
```

```
Qy    241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300
          || ||| |||| ||||||:|:|||||||  |||| |||:||||||| |||| |||:||||
Db    241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300

Qy    301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360
          ||||||||||||||||||||||:||| : ||||||||:|||:||||||||:||| |||:
Db    301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360

Qy    361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
          |||||||| |||||||||:|| ||  |||||||| | |||||||||||||| ||||||||
Db    361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420

Qy    421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480
          | |||||||||||||| ||||: ||||||:|||||| | | |||||||| :| | ||
Db    421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAA-SPSEI 479

Qy    481 PFQLRR 486
          ||:  |
Db    480 PFRPMR 485
```

RESULT 3
US-09-971-536-69
; Sequence 69, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christensson, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; TITLE OF INVENTION: Using Them
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1463
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-69

Query Match 6.1%; Score 147; DB 9; Length 1463;
Best Local Similarity 20.8%; Pred. No. 0.0025;
Matches 104; Conservative 64; Mismatches 241; Indels 92; Gaps 12;

```
Qy      1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
          :: | :|      |  :   ||||:  :|  :|:: :   :::     :|     | : |
Db    714 INANSAADVTASQAKATSAAASDAASYASEAQSIAGSHADNME-----IKSLASDAEKQS 768

Qy     61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATI-----AETFAKAEKFDRLATT 115
            |     : || : :    | :   |:   ||  : |       |   : |  :|  |:
Db    769 QIALAASKSAAASSSAAASAAIV--ASSAASEASSAAAAVSNADASANSAAAAYDSYASE 826

Qy    116 ASSA--------FENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVG 167
          ||:|         :    |||:|    |                     | ||| |
Db    827 ASAASAANDSSGYATASFAASSAAAAM---------------------SAALSTAQVAAK 865

Qy    168 TKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVR--T 225
            : | |                 | ||:    | :  :|        |    || ::  |
```

```
Db    866 VAVSDAAAAG-----------SAAAVASAAQSDSKNKQATAATARSQALDDLNKIKSLT 913

Qy    226 VLAPALASRPSVQG----------AVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGA 274

Db    914 DYASGASSSASEAGQASTATSAYASAASSSASEAGSYAHQAGSSASDAVGQSGSAAQHAS 973

Qy    275 FVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRS 334

Db    974 TAASAASSYPKDSGIQSLASQAASEAAKASSNASAATSA----AAVGFSAASDASEQAKT 1029

Qy    335 LVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTA 394

Db   1030 AASADVVASSAASTANSNASAAA----------SATKAGDSKAA---------AGFSSA 1069

Qy    395 GLATDPAVKKAESFIQDKVKSTAS--STTSYVADQTVKLAKTVKDMSGEAISSTGASLRS 452

Db   1070 ASAAASSAKGAEAVASEAASAAASDDSVASSAASAAAGFDKAASAAEGAASSAASAAASS 1129

Qy    453 TVNNLRHRSAPEADIEEGGIS 473

Db   1130 AAAQGTRGGASSSASEAGQAS 1150


RESULT 4
US-09-813-214A-9
; Sequence 9, Application US/09813214A
; Patent No. US20020177200A1
; GENERAL INFORMATION:
;  APPLICANT: Tucker, Kenneth
;  APPLICANT:  Plosila, Laura
;  TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE, GEN
;  TITLE OF INVENTION:  SEQUENCE AND USES THEREOF
;  FILE REFERENCE: 7969-089-999
;  CURRENT APPLICATION NUMBER: US/09/813,214A
;  CURRENT FILING DATE:  2000-03-20
;  PRIOR APPLICATION NUMBER: 08/968,685
;  PRIOR FILING DATE: 1997-11-12
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
;   LENGTH: 2122
;   TYPE: PRT
;   ORGANISM: Moraxella catarrhalis
US-09-813-214A-9

  Query Match             6.1%;  Score 146.5;  DB 9;  Length 2122;
  Best Local Similarity   22.2%;  Pred. No. 0.0045;
  Matches  111;  Conservative   69;  Mismatches  181;  Indels  139;  Gaps   23;

Qy     32 RSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDALVETR---ANR 88

Db    377 KTLTITGGAQTSALTD---------HNIGVVQNGDGLKVQLAETLTSLKMVTTENLTANE 427

Qy     89 LYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPL 148

Db    428 KVTVGKT---------------RL-TTDKIGFTND----MNGIDESKPYLDKDTGIHAGG 467

Qy    149 KPLTPLISGAL-----------------SGAMDQVGTKMMDRARGDLHYLSTSPDKLHDA 191

Db    468 QKITKLTAGVVDDDAATYGQLKKVNQTAESALQTFTVKKVDKNGND----------ANDS 517

Qy    192 MAVSVKRHSPALGRQVVDM------GIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGV 245

Db    518 KIITVGKNNKPDGTQVNTLKLKGENGVDVTT------------------ETNGTVTFGL 558

Qy    246 STAGGL-VANAGFGDRMLSVQ---SRDQLRGGAFVLGMKD---KEPKAALSEET----DW 294

Db    559 NQNNGLTVGNSTLNNDGLSVKNTNSNKQIQVGADGITFTDISNSKPGAGIENTTRITRDG 618

Qy    295 LDAYKAIKSASYSGAALNAGKRMAGLPLDVATD---GLKAVRSLVSATSLTKNGLALAGG 351

Db    619 I------------GFANNTGSLDANKPRLTPTGINAGGKELTNVQSAINPATNG------ 660


Qy    352 YAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAV-KKAESFIQ 410

Db    661 -GQLDFMNRLSTANTEKSGSAATIKDLYNLSQVPLTFAG------DTGPNVTKKLGEILK 713

Qy    411 DK-VKSTASSTT----SYVADQ-----TVKLAKTVKDMSGEAISSTGASLRSTVNNLRHR 460

Db    714 VKGGKTTADDLTKNNIGVVADSTDNSLTVKLAKTLSDLDAVNTKTLTASDKVTVDSGNN- 772

Qy    461 SAPEADIEEGGISAFSRSET 480

Db    773 ---TAKLQNGDLT-FSKQNT 788


RESULT 5
US-09-932-183A-2
; Sequence 2, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
;  APPLICANT: Estell, David A.
;  TITLE OF INVENTION: Proteases From Gram-Positive Organisms
;  FILE REFERENCE: GC394C1-US
;  CURRENT APPLICATION NUMBER: US/09/932,183A
;  CURRENT FILING DATE:  2001-08-17
;  PRIOR APPLICATION NUMBER: US 09/308,375
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: PCT/US98/18828
;  PRIOR FILING DATE: 1998-09-08
;  PRIOR APPLICATION NUMBER: EP9719636.4
;  PRIOR FILING DATE: 1997-09-15
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2285
;   TYPE: PRT
;   ORGANISM: Bacillius subtilis
US-09-932-183A-2

  Query Match             5.8%;  Score 140;  DB 10;  Length 2285;
  Best Local Similarity   20.5%;  Pred. No. 0.016;
  Matches  128;  Conservative   80;  Mismatches  224;  Indels  192;  Gaps   26;

Qy     15 MESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSV-IGSPDERDAALA 73

Db    251 IELYQRQAQVNVQNLNTRYGSSMGSSNRQAVQDYLN----AVNSLNVSTGSNNIRSQIQS 306

Qy     74 HNEQIDAL---VETRANRLYSEGETPATIAETF-----------------------AKA 106

Db    307 LNMQFRELASNAQTAANQASSFG---AELTQTFKSMSTYLISGSLFYGAISGLKEMVSQA 363

Qy    107 EKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQV 166

Db    364 IEIDTLMTNIRRVMNEPDYKYNELL---QESIDLGDTLSNKITDILQMTGDFGRMGFDES 420

Qy    167 GTKMMDRARGDLHYLS-TSPDKLHDAMAVSVKRHSPALGRQV--------VDMGIAVQTF 217

Db    421 ELSTLTKTAQVLQNVSDLTPDDTVNTLTAAMLNFNIAANDSISIADKLNEVDNNYAVTTL 480

Qy    218 SALNVVRTVLAPALASRPSVQGAVDFGVSTA--------GGLVAN------AGFGDRMLS 263

Db    481 DLANSIRK--AGSTASTFGVELNDLIGYTTAIASTTRESGNIVGNSLKTIFARIGNNQSS 538

Qy    264 VQSRDQLRGGAFVLGMKDKEPKAALSE---ETDWL-DAYKAIKSASYSG---------- 308

Db    539 IKALEQIGISVKTAGGEAKSASDLISEVAGKWDTLSDAQKQNTSIGVAGIYQLSRFNAMM 598

Qy    309 ----AALNAGK---------------------------------------------- 315

Db    599 NNFSIAQNAAKTAANSTGSAWSEQQKYADSLQARVNKLQNNFTEFAIAASDAFISDGLIE 658

Qy    316 --RMAGLPLDVATDGLKAV---RSLVSATS-----LTKNGLALAGG-YAGVSKL--QKMA 362

Db    659 FTQAAGSLLNASTGVIKSVGFLPPLLAAVSTATLLLSKNTRTLASSLILGTRAMGQETLA 718
```

```
Qy    363 TKNITDSATKAAVSQ-----------LSNLVGSVGVFAGWTTAGLATDPA-VKKA----- 405
Db    719 TAGLEAGMTRAAVASRVLKTALRGLLVSTLVGGAFAALGWALESLISSFAEAKKAKDDFE 778

Qy    406 -----------------ESFIQ-----DKVKSTASSTTSYVADQ---TVKLAKT----VK 436
Db    779 QSQQTNVEAITTNKDSTDKLIQQYKELQKVKESRSLTSDEEQEYLQVTQQLAQTFPALVK 838

Qy    437 --DMSGEAISSTGASLRSTVNNLR 458
Db    839 GYDSQGNAILKTNKELEKAIENTK 862
```

RESULT 6
US-09-738-626-6239
; Sequence 6239, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6239
; LENGTH: 2993
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6239

Query Match 5.7%; Score 138; DB 9; Length 2993;
Best Local Similarity 21.7%; Pred. No. 0.034;
Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;

```
Qy      7 AQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPD 66
Db   1025 AHTAPDVLVGRAWPAVFAAVKSAVIPGTDSASVVEGMLSLVHLEHHI-------VLKSDV 1077

Qy     67 ERDAALAHNEQIDALVETRANRLY--------SEGETPATIAETFA-KAEKFDRLATTAS 117
Db   1078 PTDGALKVSATADEVVDTDLGRLVIVRAEIADAEGNLIATLAERFAIRGRKGNAVARTNT 1137

Qy    118 SAFE---NTPFAAASVLQYMQP------AINKGDWLATPLKPLTPL----ISGALSGAMD 164
Db   1138 SALPTTVDTPRSARAVATVVAPESMRPFAVISGD--------RNPIHVSDVAASLAGLPG 1189

Qy    165 QVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRH-----SPALGRQVVDMGI---AVQT 216
Db   1190 VIVHGMWTSAIGELIAGAAFNDEQIQTPAAKVVEYTATMLAPVLPGEEIEFSVERSAVDN 1249

Qy    217 FSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVA--NAGFGDRMLSVQSRDQLRGGA 274
Db   1250 RPGMGEVRTVTA-------TVNGNL---VLTATAVVAAPSTFYAFPGQGIQSQG------ 1293

Qy    275 FVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVR- 333
Db   1294 --MGMEARRNSQAARAIWDRADAHTRNK------LGFSIVEIVENNPREVTVAGEKFFHP 1345
```

```
Qy    334 -SLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWT 392
Db   1346 DGVLYLTQFTQVGMATL-GVAQIAEMREAHALNQRAYFAGHSVGEYNALAAYAGVLSLES 1404

Qy    393 T------AGLATDPAVKKAESFIQ---------DKVKSTASSTTSYVADQTVKLAKTVKD 437
Db   1405 VLEIVYRRGLTMHRLVDRDENGLSNYALAALRPNKMGLTADNVFDYVA--------SVSE 1456

Qy    438 MSGE--------------AISSTGASLRSTVNNLRHRSAPEADIEE--GGISAF---SRS 478
Db   1457 ASGEFLEIVNYNLAGLQYAVAGTQAGL----------AALRADVENRAPGQRAFILIPGI 1506

Qy    479 ETPFQLRRL 487
Db   1507 DVPFHSSKL 1515
```

RESULT 7
US-09-379-931-7
; Sequence 7, Application US/09379931
; Patent No. US20020009792A1
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. US20020009792A1ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/09/379,931
; CURRENT FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614,377
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: US 08/194,290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895,367
; PRIOR FILING DATE: 1992-06-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-379-931-7

Query Match 5.5%; Score 132; DB 10; Length 1026;
Best Local Similarity 21.0%; Pred. No. 0.022;
Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;

```
Qy     22 SDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDAL 81
Db     42 SDAAALTNTLKLVNSTTAVAIQT------YQFFTGVAPSAAG--------------LDFL 81

Qy     82 VETRAN------RLYSEGETPATIAETFAKAEKF----DRLATTASSAFENTPFAAA--- 128
Db     82 VDSTTNTNDLNDAYYSK----------FAQENRFINFSINLATGAGAG--ATAFAAAYTG 129

Qy    129 -SVLQYMQPAINK--GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSP 185
Db    130 VSYAQTVATAYDKIIGNAVAT------------AAGVDVAAAVAFLSRQANIDYLT--- 173

Qy    186 DKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVL---APALASRPSVQGAVD 242
Db    174 --------AFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAAMINDLSD 225

Qy    243 FGVST--AGGL-----VANAGFGDRMLSVQS-RDQLRG------------GAFVLGMKD 281
Db    226 GALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVAGEVAGAATLTVGD 285

Qy    282 KEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSL 341
Db    286 TLSGGAGTDVLNWVQA----------AAVTA------LPTGVTISGIETM-NVTSGAAI 327

Qy    342 TKNGLALAGGYAGVSKL---------------QKMATKNITDSATKAAVSQLSNL-VGS 384
```

```
Db    328 TLN---TSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGRANVTVAS 384

Qy    385 VGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAIS 444

Db    385 TGVTSGTTTVG-----ANSAASGTVSVSVANSSTTTTGAIA-VTGGTAVTVAQTAGNAVN 438

Qy    445 ST 446

Db    439 TT 440
```

RESULT 8
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match 5.4%; Score 130.5; DB 10; Length 2368;
Best Local Similarity 20.2%; Pred. No. 0.093;
Matches 99; Conservative 64; Mismatches 223; Indels 105; Gaps 17;

```
Qy      4 NQSAQQPPGVAMESFRTAS-------DASLASSSVRSVSTTSCRDLQAITDYLKHHVFAA 56

Db    702 NASLQDEKDVANDKIGKIETKAIKDIDAATTNAQVEAIKTKAINDINQTAPATTAKAAAL 761

Qy     57 HRFSVIGSPDERDAAL---AHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLA 113

Db    762 EEFDEVVQAQIDQAPLNPDTTNEEVAEAIE----RINAAKVSGVKAIEATTTAQDLERVK 817

Qy    114 TTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDR 173

Db    818 NEEISKIEN----ITDSTQTKMDAYNEVKQAATARKTQNATVSNATN---EEV------- 863

Qy    174 ARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALAS 233

Db    864 AEADAAVEAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAIQT---------------- 906

Qy    234 RPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETD 293

Db    907 QAKVKPAADTEVENA--------YNTRKQEIQNSN---------ASTTEEKQAAYTE--- 946

Qy    294 WLDAYKAIKSASYSGAALNAGKRMAGLPLDVAT---DGLKAVRSLVSATSLTKNGLALAG 350

Db    947 -LDTKKQEARTNLDAANTNS---------DVTTAKDNGIAAINQVQAATTKKSDAKAEIA 996

Qy    351 GYAGVSKL-----------QKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTT-----A 394

Db    997 QKASERKTAIEAMNDSTTEEQQAAKDKVDQAVVTANADIDNAAANTDVDNAKTTNEATIA 1056

Qy    395 GLATDPAVK-KAESFIQDKVKST---------ASSTTSYVADQTVKLAKTVKDMSGEAIS 444

Db   1057 AITPDANVKPTAKQAIADKVQAQETAIDANNGATTEEKAAAKQQVQTEKTTADTAIDG-A 1115

Qy    445 STGASLRSTVN 455

Db   1116 HTNAEVEAAKN 1126
```

RESULT 9
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12389
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12389

Query Match 5.4%; Score 130.5; DB 10; Length 2368;
Best Local Similarity 20.2%; Pred. No. 0.093;
Matches 99; Conservative 64; Mismatches 223; Indels 105; Gaps 17;

```
Qy      4 NQSAQQPPGVAMESFRTAS-------DASLASSSVRSVSTTSCRDLQAITDYLKHHVFAA 56

Db    702 NASLQDEKDVANDKIGKIETKAIKDIDAATTNAQVEAIKTKAINDINQTAPATTAKAAAL 761

Qy     57 HRFSVIGSPDERDAAL---AHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLA 113

Db    762 EEFDEVVQAQIDQAPLNPDTTNEEVAEAIE----RINAAKVSGVKAIEATTTAQDLERVK 817

Qy    114 TTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDR 173
```

```
Db    818 NEEISKIEN----ITDSTQTKMDAYNEVKQAATARKTQNATVSNATN---EEV------- 863

Qy    174 ARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALAS 233

Db    864 AEADAAVEAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAIQT---------------- 906

Qy    234 RPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETD 293

Db    907 QAKVKPAADTEVENA--------YNTRKQEIQNSN---------ASTTEEKQAAYTE--- 946

Qy    294 WLDAYKAIKSASYSGAALNAGKRMAGLPLDVAT---DGLKAVRSLVSATSLTKNGLALAG 350

Db    947 -LDTKKQEARTNLDAANTNS---------DVTTAKDNGIAAINQVQAATTKKSDAKAEIA 996

Qy    351 GYAGVSKL-----------QKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTT-----A 394

Db    997 QKASERKTAIEAMNDSTTEEQQAAKDKVDQAVVTANADIDNAAANTDVDNAKTTNEATIA 1056

Qy    395 GLATDPAVK-KAESFIQDKVKST---------ASSTTSYVADQTVKLAKTVKDMSGEAIS 444

Db   1057 AITPDANVKPTAKQAIADKVQAQETAIDANNGATTEEKAAAKQQVQTEKTTADTAIDG-A 1115

Qy    445 STGASLRSTVN 455

Db   1116 HTNAEVEAAKN 1126


RESULT 10
US-09-815-242-5815
; Sequence 5815, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5815
;  LENGTH: 2076
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-5815

  Query Match             5.4%;  Score 129.5;  DB 10;  Length 2076;
  Best Local Similarity   19.3%;  Pred. No. 0.093;
  Matches  106;  Conservative   91;  Mismatches  201;  Indels  151;  Gaps   22;

Qy     20 TASDASL----ASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHN 75

Db    930 TAANTNILNANTNADVEQVKTNAIQGIQAITPATK----------VKTDAKNA----- 972

Qy     76 EQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQ 135

Db    973 --IDKSAETQHNTIFNNND--ATLEEQQAAQQLLDQAVATAKQNIN-----AADTNQEVA 1023

Qy    136 PAINKGDWLATPLKPLTPLISGALSGAMDQVGTK----MMDRARGDLHYLSTSPDKLHDA 191

Db   1024 QAKDQGTQNIVVIQPAT----------QVKTDTRNVVNDKAREAITNINATTGATREE 1071

Qy    192 MAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRT-------VLAPALASRPSVQGA---- 240

Db   1072 KQEAINRVNTLKNRALTDIGV-TSTTAMVNSIRDDAVNQIGAVQPHVTKKQTATGVLNDL 1130

Qy    241 -------------------------VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAF 275

Db   1131 ATAKKQEINQNTNATTEEKQVALNQVDQELATA---INNINQADTNAEVDQAQQLGTKAI 1187

Qy    276 ------------VLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMA----- 318

Db   1188 NAIQPNIVKKPAALAQINQHYNAKLAEINATPDATNDEKNAAIN--TLNQDRQQAIESIK 1245

Qy    319 ----GLPLD----VATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMA--TKNITD 368

Db   1246 QANTNAEVDQAATVAENNIDAVQVDVVKKQAARDKIT-----AEVAKRIEAVKQTPNATD 1300

Qy    369 SATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKA-ESFIQDKVKSTASSTTSYV--- 424

Db   1301 EEKQAAVNQINQL----------------KDQAINQINQNQTNDQVDTTTNQAVNAIDNV 1344

Qy    425 -ADQTVK------LAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSR 477

Db   1345 EAEVVIKPKAIADIEKAVKEKQQQIDN----SLDSTDNEKEVASQALAKEKEKALAAIDQ 1400

Qy    478 SETPFQLRR 486

Db   1401 AQTNSQVNQ 1409


RESULT 11
US-09-815-242-12913
; Sequence 12913, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12913
;   LENGTH: 2186
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12913

  Query Match             5.4%;  Score 129.5;  DB 10;  Length 2186;
  Best Local Similarity   19.3%;  Pred. No. 0.1;
  Matches  106;  Conservative   91;  Mismatches  201;  Indels  151;  Gaps   22;

Qy      20 TASDASL----ASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHN 75
           ||:: ::      :: |  | | : : :||||   |          | ::|
Db     930 TAANTNILNANTNADVEQVKTNAIQGIQAITPATK------------VKTDAKNA----- 972

Qy      76 EQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQ 135
             ||   ||: | :::  :  ||: |  |  :  |:   ||          ||    | :
Db     973 --IDKSAETQHNTIFNNND--ATLEEQQAAQQLLDQAVATAKQNIN-----AADTNQEVA 1023

Qy     136 PAINKGDWLATPLKPLTPLISGALSGAMDQVGTK----MMDRARGDLHYLSTSPDKLHDA 191
           | ::|      ::| |             || |      : |:||  :  :: :     :
Db    1024 QAKDQGTQNIVVIQPAT-----------QVKTDTRNVVNDKAREAITNINATTGATREE 1071

Qy     192 MAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRT-------VLAPALASRPSVQGA---- 240
             :: | :    | : |:|:   | : :| :|        : | :  :  :  |
Db    1072 KQEAINRVNTLKNRALTDIGV-TSTTAMVNSIRDDAVNQIGAVQPHVTKKQTATGVLNDL 1130

Qy     241 -------------------------VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAF 275
                                     ||  ::||   : |    |    |    ||   |
Db    1131 ATAKKQEINQNTNATTEEKQVALNQVDQELATA---INNINQADTNAEVDQAQQLGTKAI 1187

Qy     276 ------------VLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMA----- 318
                       |   ::   | |:|      ||     |:|: :   ||  :: |
Db    1188 NAIQPNIVKKPAALAQINQHYNAKLAEINATPDATNDEKNAAIN--TLNQDRQQAIESIK 1245

Qy     319 ----GLPLD----VATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMA--TKNITD 368
                 :|    || : : ||:  |       :: :       | |:|  :    | | ||
Db    1246 QANTNAEVDQAATVAENNIDAVQVDVVKKQAARDKIT-----AEVAKRIEAVKQTPNATD 1300

Qy     369 SATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKA-ESFIQDKVKSTASSTTSYV--- 424
           :|||:|:: |                 .| |: :  ::   |:| :| :   : :
Db    1301 EEKQAAVNQINQL----------------KDQAINQINQNQTNDQVDTTTNQAVNAIDNV 1344

Qy     425 -ADQTVK------LAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSR 477
            |:  :|      : | ||:   :  :    || || |    |   | :|  ::|  :
Db    1345 EAEVVIKPKAIADIEKAVKEKQQQIDN----SLDSTDNEKEVASQALAKEKEKALAAIDQ 1400

Qy     478 SETPFQLRR 486
           ::|  |: :
Db    1401 AQTNSQVNQ 1409


RESULT 12
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12713
;   LENGTH: 2344
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12713

  Query Match             5.4%;  Score 129;  DB 10;  Length 2344;
  Best Local Similarity   17.7%;  Pred. No. 0.12;
  Matches   86;  Conservative   97;  Mismatches  222;  Indels   82;  Gaps   12;

Qy      16 ESFRTASDASLASSSVRSVST-TSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAH 74
             :|  |:|  | ::|   | || ||  |  :|:          :   |: ||  | |:
Db    1421 DSQSTSSYTSQSTSQSESTSTSTSLSDSTSISKSTSQSGSTSTSASLSGSESESDSQSIS 1480

Qy      75 NEQIDALVETRANRLYSEGET-------PATIAETFAKAEKFDRLATTASSAFENTPFAA 127
                ::  |: :  |      |         :|:    | | : |  :|: | :   :  ::
Db    1481 TSTSESKSESTSTSLSDSTSTSNSGSASTSTLLSNSASASESDSSSTSLSDSTSASMQSS 1540

Qy     128 ASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVG------------------- 167
             |  |    :::      :| :: :: : | ::| :  :  |
Db    1541 ESDSQSTSTSLSNSQSTSTSIR-MSTIASESVSESTSESGSTSESTSESDSTSTSLSDSQ 1599

Qy     168 -TKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTV 226
              |      | |      ::: |    : : |    :  |  |  | ::  | ::::
Db    1600 STSRSTSASGSASTSTSTSDSRSTSASTSTSMRTSTLDSQ--SMSLSTSTSTSVSD---- 1653

Qy     227 LAPALASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKA 286
             : :|:   |   :     ||:| : |:    |   :  |  ::                |
Db    1654 -STSLSDSVSDSTSDSTSTSTSGSMSASISLSDSTSTSTSASEVM-------------SA 1699

Qy     287 ALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGL 346
           ::|:      ::    :| | | :  :: | |:|    |:  |      ||  :||| |
Db    1700 SISDSQSMSESVNDSESVSESNSESDS-KSMSG-STSVSDSG-----SLSVSTSLRK--- 1749

Qy     347 ALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAE 406
                           ::::::| : :    :|: | :         :: |:   : : :|
Db    1750 ---------------SESVSESISLSGSQSMSDSVSTS------DSSSLSVSTSQRSSE 1787

Qy     407 SFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEAD 466
           |  :    | : ||:   :  |     |   :||    |  :||  :::      |  ::|
Db    1788 SVSESDSLSDSKSTSGSTSTSTSGSLSTSTSLSGSESVSESSSLSDSISMSDSTSTSDSD 1847

Qy     467 IEEGGIS 473
             | ||
Db    1848 SLSGSIS 1854


RESULT 13
US-09-815-242-5816
; Sequence 5816, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
```

```
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5816
;   LENGTH: 2478
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-5816

  Query Match             5.3%;  Score 127.5;  DB 10;  Length 2478;
  Best Local Similarity   19.1%;  Pred. No. 0.17;
  Matches  106;  Conservative   70;  Mismatches  235;  Indels  143;  Gaps   18;

Qy       4 NQSAQQPPGVAMESFRTAS-------DASLASSSVRSVSTTSCRDLQAITDYLKHHVFAA 56
           | | |    || :              ||:  :: | :: | :  |:   |         |
Db     706 NASLQDEKDVANDKIGKIETKAIKDIDAATTNAQVEAIKTKAINDINQTTPATTAKAAAL 765

Qy      57 HRFSVIGSPDERDAAL---AHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLA 113
             |  :       | |     ||::   :|    |: :   :      |    |:  :|:
Db     766 EEFDEVVQAQIDQAPLNPDTTNEEVAEAIE----RINAAKVSGVKAIEATTTAQDLERVK 821

Qy     114 TTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDR 173
             |  ||          |    | |:    ||  |     :| | :   ::|
Db     822 NEEISKIEN----ITDSTQTKMDAYNEVKQAATARKAQNATVSNATN---EEV------- 867

Qy     174 ARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALAS 233
           |  |    :      |||    |  |:       :|:|   |:||
Db     868 AEADAAVDAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAIQT---------------- 910

Qy     234 RPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETD 293
           :  |: | |  |  |       :  |   :|: :              :| :|| :|
Db     911 QAKVKPAADTEVENA--------YNTRKQEIQNSN---------ASTTEEKQAAYTE--- 950

Qy     294 WLDAYKAIKSASYSGAALNAGKRMAGLPLDVAT---DGLKAVRSLVSATSLTKNGLALAG 350
           ||  |     :   |  |:          || |   : : |:  : :||:   :  |
Db     951 -LDTKKQEARTNLDAANTNS---------DVTTAKDNSIAAINQVQAATTKKSDAKAEIA 1000

Qy     351 GYAGVSKL-----------QKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTT-----A 394
           |   |              :: | |:  | |   | : : |   :  |    ||     |
Db    1001 QKASERKTAIEAMNDSTTEEQQAAKDKVDQAVVTANADIDNAAANNDVDNAKTTNEATIA 1060

Qy     395 GLATDPAVK-KAESFIQDKVKS-----------------------TASSTTSYVAD--- 426
           :  |  ||  |:  | |||::   .                    |  :|     |
Db    1061 AITPDANVKPAAKQAIADKVQAQETAIDGNNGSTTEEKAAAKQQVQTEKTTADAAIDAAH 1120

Qy     427 ----------------QTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPE-----A 465
                           : :: | | || : |||::    :: :   :  :| |     |
Db    1121 TNAEVEAAKKAAIAKIEAIQPATTTKDNAKEAIATKANERKTAIAQTQDITAEEIAAANA 1180

Qy     466 DIEEGGISAFSRSE 479
           |::     | |  |
Db    1181 DVDNAVTQANSNIE 1194

RESULT 14
US-09-815-242-12967
; Sequence 12967, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12967
;   LENGTH: 2478
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12967

  Query Match             5.3%;  Score 127.5;  DB 10;  Length 2478;
  Best Local Similarity   19.1%;  Pred. No. 0.17;
  Matches  106;  Conservative   70;  Mismatches  235;  Indels  143;  Gaps   18;

Qy       4 NQSAQQPPGVAMESFRTAS-------DASLASSSVRSVSTTSCRDLQAITDYLKHHVFAA 56
           | | |    || :              ||:  :: | :: | :  |:   |         |
Db     706 NASLQDEKDVANDKIGKIETKAIKDIDAATTNAQVEAIKTKAINDINQTTPATTAKAAAL 765

Qy      57 HRFSVIGSPDERDAAL---AHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLA 113
             |  :       | |     ||::   :|    |: :   :      |    |:  :|:
Db     766 EEFDEVVQAQIDQAPLNPDTTNEEVAEAIE----RINAAKVSGVKAIEATTTAQDLERVK 821

Qy     114 TTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDR 173
             |  ||          |    | |:    ||  |     :| | :   ::|
Db     822 NEEISKIEN----ITDSTQTKMDAYNEVKQAATARKAQNATVSNATN---EEV------- 867

Qy     174 ARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALAS 233
           |  |    :      |||    |  |:       :|:|   |:||
Db     868 AEADAAVDAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAIQT---------------- 910

Qy     234 RPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETD 293
           :  |: | |  |  |       :  |   :|: :              :| :|| :|
Db     911 QAKVKPAADTEVENA--------YNTRKQEIQNSN---------ASTTEEKQAAYTE--- 950

Qy     294 WLDAYKAIKSASYSGAALNAGKRMAGLPLDVAT---DGLKAVRSLVSATSLTKNGLALAG 350
           ||  |     :   |  |:          || |   : : |:  : :||:   :  |
Db     951 -LDTKKQEARTNLDAANTNS---------DVTTAKDNSIAAINQVQAATTKKSDAKAEIA 1000

Qy     351 GYAGVSKL-----------QKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTT-----A 394
           |   |              :: | |:  | |   | : : |   :  |    ||     |
Db    1001 QKASERKTAIEAMNDSTTEEQQAAKDKVDQAVVTANADIDNAAANNDVDNAKTTNEATIA 1060

Qy     395 GLATDPAVK-KAESFIQDKVKS-----------------------TASSTTSYVAD--- 426
           :  |  ||  |:  | |||::                         |  :|     |
```

```
Db   1061 AITPDANVKPAAKQAIADKVQAQETAIDGNNGSTTEEKAAAKQQVQTEKTTADAAIDAAH 1120

Qy    427 ----------------QTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPE-----A 465
                          : :: | | || : |||::     :: :   :  :| |     |
Db   1121 TNAEVEAAKKAAIAKIEAIQPATTTKDNAKEAIATKANERKTAIAQTQDITAEEIAAANA 1180

Qy    466 DIEEGGISAFSRSE 479
          |::     | |  |
Db   1181 DVDNAVTQANSNIE 1194
```

```
RESULT 15
US-09-137-531-15
; Sequence 15, Application US/09137531
; Patent No. US20020048816A1
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/137,531
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/682,517
;      FILING DATE:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1222 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-137-531-15

  Query Match             5.2%;  Score 125.5;  DB 10;  Length 1222;
  Best Local Similarity   20.3%;  Pred. No. 0.091;
  Matches  117;  Conservative   73;  Mismatches  219;  Indels  167;  Gaps   25;

Qy      3 INQSAQQPPGVAME-----------------------SFRTASDASLASSSVRSVSTT 37
          :||| ::  | :::                       |:   :|:|:|:       |
Db    675 VNQSGKKVTGTSIKKATYTIYNTGANDIKVDNQVISPNRSYTVTYEATLSST---GTVIT 731

Qy     38 SCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDALVET------------ 84
            ::|:  :   |          :  : | :| |    |      |
Db    732 PAKNLEVTSVDGKTTAVKVIATGIAVNTDGKDYAFTAKEATATFTATNEVPNSYTGVATQ 791

Qy     85 --------RANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQY--- 133
                  :| ::  |: |   |     |  |:       :  |    : |  : ||
Db    792 FNTADSGSNSNSIWFAGKNPVKYAGVSGKTYKY--FGANGNEVFGEAAWEAL-LTQYATE 848

Qy    134 ------------------MQPAINKGDWLATPLKPLTPL--ISGALS------GAMD--- 164
                           :  |:|       |: | ||    :|||:      | :|
Db    849 GQKVTISYNVDGDTVTFKVISAVNSSTEAIKPVAPTTPAAPTTGALTLTPAAGGLVDLTT 908

Qy    165 QVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVK-RHSPALGRQVVDMGIAVQTFSALNVV 223
            |  :  |  ||:  :|:      |   ||:|  : :|   :|: |      |:
Db    909 ATNTLGISLADADLNVSATTV----DTATVSLKDSANNSLSLTLVETGANTGVFAT---- 960

Qy    224 RTVLAPALASRPSVQGAVDFGVSTAGGLVAN-------AGFGDRMLSVQSRDQLRGG--- 273
           || |  |:|          |||  |          ||  : : :  :  :  |
Db    961 -TVQAGTLSS-----------LTAGTLTVTYADAKNAAGVAENITASVTLKKTTGAITS 1007

Qy    274 -AFVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAAL---NAGKRMAGLPLDVATDGL 329
           |  |:    | || : |             |:  |  |   ||| ::  |       |
Db   1008 DTFTQGVL---PSAATAAEYTSKSIAADYTFATGEGFTLNIDNAGAQVINLA------GK 1058

Qy    330 KAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITD------SATKAAVSQLSNLV- 382
          |  : :  | : |  | |   |         |:  |: |          | ::|:|: | |
```

```
Db   1059 KGAQGVADAINATFAGTATVSG-------DKVVIKSATTGVGSEVEVTFSSVNQVLNAVV 1111

Qy    383 -GSVGVFAGW-TTAGLATDPAVKKAESFIQDKVKST----------------ASSTTSY 423
           |   | ||   |        |:   |  : | |: |                  |::| :
Db   1112 NGKDQVVAGTAATKAFTITTALSVGEKVVIDGVEYTAVAFGTAPTANTFVVESAANTLAS 1171

Qy    424 VADQTVKLAKTVKDMSGE---AISSTGA--SLRSTV 454
          ||||   || |:  ::       |:|||  :| |||
Db   1172 VADQAANLAATIDTLNTADKFTASATGATITLTSTV 1207
```

Search completed: January 2, 2003, 15:22:17
Job time : 19.5087 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:19:06 ; Search time 8.49126 Seconds
(without alignments)
1084.696 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHINRRVQQPPVTATDSFRT.....IEEGGTAASPSEIPFRPMRS 486

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 117078 seqs, 18951520 residues

Total number of hits satisfying chosen parameters: 117078

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2412	100.0	486	10	US-09-825-414-7	Sequence 7, Appli
2	1869.5	77.5	487	10	US-09-825-414-66	Sequence 66, Appl
3	146	6.1	1463	9	US-09-971-536-69	Sequence 69, Appl
4	143	5.9	600	10	US-09-287-849-22	Sequence 22, Appl
5	133.5	5.5	729	10	US-09-287-849-2	Sequence 2, Appli
6	131.5	5.5	2086	10	US-09-815-242-5639	Sequence 5639, Ap
7	131.5	5.5	5795	10	US-09-815-242-12610	Sequence 12610, A
8	129.5	5.4	2344	10	US-09-815-242-12713	Sequence 12713, A
9	129	5.3	2434	10	US-09-815-242-5835	Sequence 5835, Ap
10	129	5.3	6281	10	US-09-815-242-12996	Sequence 12996, A
11	127.5	5.3	596	10	US-09-287-849-26	Sequence 26, Appl
12	127	5.3	2122	9	US-09-813-214A-9	Sequence 9, Appli
13	126.5	5.2	1026	10	US-09-379-931-7	Sequence 7, Appli
14	125	5.2	714	10	US-09-841-786-4	Sequence 4, Appli
15	125	5.2	3241	10	US-09-841-786-1	Sequence 1, Appli
16	124.5	5.2	1744	9	US-10-108-605-25	Sequence 25, Appl
17	122	5.1	5215	9	US-09-860-846-2	Sequence 2, Appli
18	122	5.1	5215	10	US-09-861-289-2	Sequence 2, Appli
19	120.5	5.0	1879	9	US-09-971-536-70	Sequence 70, Appl
20	118	4.9	2285	10	US-09-932-183A-2	Sequence 2, Appli
21	116.5	4.8	573	10	US-09-828-574-13	Sequence 13, Appl
22	116	4.8	2478	10	US-09-815-242-5816	Sequence 5816, Ap
23	116	4.8	2478	10	US-09-815-242-12967	Sequence 12967, A
24	116	4.8	2993	9	US-09-738-626-6239	Sequence 6239, Ap
25	115	4.8	342	10	US-09-815-242-11043	Sequence 11043, A
26	115	4.8	1832	9	US-10-014-717-4	Sequence 4, Appli
27	114	4.7	2368	10	US-09-815-242-5635	Sequence 5635, Ap
28	114	4.7	2368	10	US-09-815-242-12389	Sequence 12389, A
29	114	4.7	7257	9	US-10-014-717-5	Sequence 5, Appli
30	113.5	4.7	650	10	US-09-815-242-13341	Sequence 13341, A
31	113	4.7	596	9	US-10-063-547-100	Sequence 100, App
32	113	4.7	596	12	US-10-006-867-100	Sequence 100, App
33	113	4.7	596	12	US-10-052-586-310	Sequence 310, App
34	113	4.7	1488	9	US-09-738-626-5495	Sequence 5495, Ap
35	112	4.6	2076	10	US-09-815-242-5815	Sequence 5815, Ap
36	112	4.6	2186	10	US-09-815-242-12913	Sequence 12913, A
37	112	4.6	4999	9	US-09-976-059-14	Sequence 14, Appl
38	111.5	4.6	1289	9	US-09-712-363-259	Sequence 259, App
39	110	4.6	647	10	US-09-841-132-436	Sequence 436, App
40	110	4.6	792	9	US-10-025-380-1127	Sequence 1127, Ap
41	109.5	4.5	1186	10	US-09-826-752-8	Sequence 8, Appli
42	108	4.5	856	10	US-09-287-849-12	Sequence 12, Appl
43	107.5	4.5	1165	12	US-10-016-768-10	Sequence 10, Appl
44	107.5	4.5	1221	12	US-10-016-768-11	Sequence 11, Appl
45	107	4.4	1054	10	US-09-798-042-87	Sequence 87, Appl

ALIGNMENTS

RESULT 1
US-09-825-414-7
; Sequence 7, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
US-09-825-414-7

Same inV.
Same

Query Match 100.0%; Score 2412; DB 10; Length 486;
Best Local Similarity 100.0%; Pred. No. 4.3e-176;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60

Qy      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120

Qy     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180
```

```
Qy     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240

Qy     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300

Qy     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360

Qy     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420

Qy     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEIP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEIP 480

Qy     481 FRPMRS 486
           ||||||
Db     481 FRPMRS 486
```

RESULT 2
US-09-825-414-66
; Sequence 66, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 66
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. tomato
US-09-825-414-66

Query Match 77.5%; Score 1869.5; DB 10; Length 487;
Best Local Similarity 78.4%; Pred. No. 8e-135;
Matches 381; Conservative 33; Mismatches 71; Indels 1; Gaps 1;

```
Qy       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60
           ||||:  ||||  | :|||||||||||||||||||:   |:: || |||  ||||||:
Db       1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60

Qy      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120
             | | : |   || || ||:||||:||: |||||||||:|||||| |||||| | |
Db      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120

Qy     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180
             |||| ||:|||||||||||||  |||||||||||||||||||||||||||| |||||
Db     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180

Qy     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240
           || |||:|||||| |||||||:| |||:| |:||||:|| | ||||||||||||:||||
Db     181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240

Qy     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300
           || ||| |||| ||||||:|:|||||||  |||| |||:|||||| |||| |||:||||
Db     241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300

Qy     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360
           ||||||||||||||||||||:||| : ||||||||:|||:||||||||:||| |||:
Db     301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360

Qy     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420
           |||||||| |||||||||:|| ||  |||||||| | ||||||||||||| |||||||||
Db     361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420

Qy     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAA-SPSEI 479
           | |||||||||||||| |||||: |||||:|||||| | | |||||||| :| | ||
Db     421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480

Qy     480 PFRPMR 485
           ||:  |
Db     481 PFQLRR 486
```

RESULT 3
US-09-971-536-69
; Sequence 69, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christensson, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; TITLE OF INVENTION: Using Them
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1463
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-69

Query Match 6.1%; Score 146; DB 9; Length 1463;
Best Local Similarity 22.5%; Pred. No. 0.0049;
Matches 112; Conservative 72; Mismatches 219; Indels 94; Gaps 20;

```
Qy      14 ATDSFRTASDASLASSSVR---SVSSD--QQREINAIADYL------TDHVFAAHKLPPA 62
           |::: : :|:|| |:|:     | :||  :| :  | || :       | :  |:
Db     996 ASEAAKASSNASAATSAAAVGFSAASDASEQAKTAASADVVASSAASTANSNASAAASAT 1055

Qy      63 DSADGQAAVDVHNAQITALIETR-ASRLHFEGETPATIADTFAK-----AEKLDRLATTT 116
            : | :||    :|   |    : |  :  |  : |   |: |       |   |: |:
Db    1056 KAGDSKAAAGFSSAASAAASSAKGAEAVASEAASAAASDDSVASSAASAAAGFDKAASAA 1115

Qy     117 SGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATG 176
            ||  :   | ||       |  :|               |  | :  : |      :|
Db    1116 EGAASSAASAAAS------SAAAQG-------------TRGGASSSASEAG----QASTA 1152

Qy     177 DLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPA 236
             | ||:    :  | |    : | | :   || |    |  :|     | : |||
```

```
Db   1153 TSVYASAASSSASE--AGSYAHQAGSSASEA--TGHASSATSQASA-----ASSAASRYP 1203

Qy    237 VQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLE 296

Db   1204 SDSGIQSDVSIASSAASTAS------SAASAAQSEASTASSAASHASEQASIASSEDVVS 1257

Qy    297 AYKAIKSASYSGAALNAGK----RMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGF 352

Db   1258 SSAASVASSAASAASSAAKAGNSSAAGIYSHAASAAASSAKSAESQASSAASAAA----- 1312

Qy    353 AGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAV-------KKA 405

Db   1313 ---------SDDSVASSAASAALSDDAKASSAADVASSATTAAISSATSLADQSATGSTA 1363

Qy    406 ESFIQDTVKSTASSTTGYV-ADQT--------VKLAKTVKDMGGEAITHTGASLRNTVNN 456

Db   1364 GSHI---LPSTGGETTGSIPSGQTPTQTKPTQTKPTQTKPTQAGQT-TQTG-SLPQTDHA 1418

Qy    457 LRQRPAREADIEEGGTA 473

Db   1419 GRHMLPQTGDDAESGTS 1435


RESULT 4
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Alderson, Mark
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Corixa Corporation
;  TITLE OF INVENTION:  Fusion Protiens of Mycobacterium tuberculosis Antigens
;  TITLE OF INVENTION:  and Their Uses
;  FILE REFERENCE:  014058-009020US
;  CURRENT APPLICATION NUMBER: US/09/287,849
;  CURRENT FILING DATE:  1999-04-07
;  PRIOR APPLICATION NUMBER:  US 08/818,112
;  PRIOR FILING DATE:  1997-03-13
;  PRIOR APPLICATION NUMBER:  US 08/942,578
;  PRIOR FILING DATE:  1997-10-01
;  PRIOR APPLICATION NUMBER:  US 09/025,197
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 09/056,556
;  PRIOR FILING DATE:  1998-04-07
;  PRIOR APPLICATION NUMBER:  US 09/223,040
;  PRIOR FILING DATE:  1998-12-30
;  NUMBER OF SEQ ID NOS:  46
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 22
;   LENGTH: 600
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

  Query Match             5.9%;  Score 143;  DB 10;  Length 600;
  Best Local Similarity   23.0%;  Pred. No. 0.0024;
  Matches  117;  Conservative   58;  Mismatches  165;  Indels  168;  Gaps   25;

Qy      20 TASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQA---------- 69

Db      88 TAGQAELTAAQVRV----------AAAAYETAYGLT---VPPPVIAENRAELMILIATNL 134

Qy      70 ------AVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL--- 120

Db     135 LGQNTPAIAVNEAEYGEMWAQDAAAMFGYAAATATATATLLPFEEAPEM--TSAGGLLEQ 192

Qy     121 -----RATPFAMAS-LLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVG------T 168
```

```
Db    193 AAAVEEASDTAAANQLMNNVPQALQQ---LAQPTQGTTP--SSKLGGLWKTVSPHRSPIS 247

Qy    169 KMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLA 228

Db    248 NMVSMANNHMS-MTNSGVSMTNTLSSMLKGFAPAAAAQ------AVQT-AAQNGVR---- 295

Qy    229 PALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQL 288

Db    296 -AMSSLGSSLGSSGLG----GGVAANLG---RAASVGS---------------------L 326

Qy    289 SEENDWLEAYKAIKSAS--------YSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASS 340

Db    327 SVPQAWAAANQAVTPAARALPLTSLTSAAERGPGQMLGGLPVG----------------- 369

Qy    341 LTQNGLALAGGFAGV----------------GKL--------------QEMATKNITDP- 369

Db    370 --QMGARAGGGLSGVLRVPPRPYVMPHSPAAGKLDPVDAVINTTCNYGQVVAALNATDPG 427

Qy    370 ------ATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASSTTGY 423

Db    428 AAAQFNASPVAQSYLRNFLAAPPPQRAAMAAQL---QAVPGAAQYI-GLVESVAGSCNNY 483

Qy    424 VADQTVKLAKTVKDMGGEAITHTGASLR 451

Db    484 E-------LMTINYQFGDVDAH-GAMIR 503


RESULT 5
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Alderson, Mark
;  APPLICANT:  Campos-Neto, Antonio
;  APPLICANT:  Corixa Corporation
;  TITLE OF INVENTION:  Fusion Protiens of Mycobacterium tuberculosis Antigens
;  TITLE OF INVENTION:  and Their Uses
;  FILE REFERENCE:  014058-009020US
;  CURRENT APPLICATION NUMBER: US/09/287,849
;  CURRENT FILING DATE:  1999-04-07
;  PRIOR APPLICATION NUMBER:  US 08/818,112
;  PRIOR FILING DATE:  1997-03-13
;  PRIOR APPLICATION NUMBER:  US 08/942,578
;  PRIOR FILING DATE:  1997-10-01
;  PRIOR APPLICATION NUMBER:  US 09/025,197
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 09/056,556
;  PRIOR FILING DATE:  1998-04-07
;  PRIOR APPLICATION NUMBER:  US 09/223,040
;  PRIOR FILING DATE:  1998-12-30
;  NUMBER OF SEQ ID NOS:  46
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 729
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-2

  Query Match             5.5%;  Score 133.5;  DB 10;  Length 729;
  Best Local Similarity   21.0%;  Pred. No. 0.017;
  Matches  116;  Conservative   70;  Mismatches  178;  Indels  189;  Gaps   25;

Qy      20 TASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQA---------- 69

Db     221 TAGQAELTAAQVRV----------AAAAYETAYGLT---VPPPVIAENRAELMILIATNL 267

Qy      70 ------AVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL--- 120
```

```
Db    268 LGQNTPAIAVNEAEYGEMWAQDAAAMFGYAAATATATATLLPFEEAPEM--TSAGGLLEQ 325

Qy    121 -----RATPFAMAS-LLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVG------T 168

Db    326 AAAVEEASDTAAANQLMNNVPQALQQ---LAQPTQGTTP--SSKLGGLWKTVSPHRSPIS 380

Qy    169 KMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLA 228

Db    381 NMVSMANNHMS-MTNSGVSMTNTLSSMLKGFAPAAARQ------AVQT-AAQNGVR---- 428

Qy    229 PALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQL 288

Db    429 -AMSSLGSSLGSSGLG----GGVAANLG---RAASVGS--------------------L 459

Qy    289 SEENDWLEAYKAIKSAS--------YSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASS 340

Db    460 SVPQAWAAANQAVTPAARALPLTSLTSAAERGPGQMLGGLPVG----------------- 502

Qy    341 LTQNGLALAGGFAGVGKL--------QEMATKNITDPA---------------TKAAVSQ 377

Db    503 --QMGARAGGGLSGVLRVPPRPYVMPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQ 560

Qy    378 L--------TNLAGSAAVFAGWTTAALTTDP----------AVKKAESFIQDTVKSTA 417

Db    561 VGPQVVNINTKLGYNNAVGAG---TGIVIDPNGVVLTNNHVIAGATDINAFSVGSGQTYG 617

Qy    418 SSTTGYVADQTVKLAK--------TVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEE 469

Db    618 VDVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMGNS---------------- 661

Qy    470 GGTAASPSEIPFR 482

Db    662 GGQGGTPRAVPGR 674
```

RESULT 6
US-09-815-242-5639
; Sequence 5639, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5639
; LENGTH: 2086
; TYPE: PRT
; ORGANISM: Staphylococcus aureus

US-09-815-242-5639

Query Match 5.5%; Score 131.5; DB 10; Length 2086;
Best Local Similarity 21.6%; Pred. No. 0.1;
Matches 124; Conservative 70; Mismatches 226; Indels 153; Gaps 26;

```
Qy      3 INRRVQ-QPPVTATDSFRTASDASLASSSVRSV----------SSDQQREI 42

Db    185 INNAVQVANKRTATIKNGTAMPTNLAGGSTTTIPVTVTYNDGSTEEVQESIFTKADKREL 244

Qy     43 NAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQ---ITALIETRASRLHFEGETPATI 99

Db    245 ITAKNHLDDPVSTDGKKPGTITQYNNA---IHNAQQQINTAKTEAQ-QVINNERATPQQV 300

Qy    100 ADTFAKAEKLDRLATTTSGALRATPFAMASLLQ----------YMQPAINKGDWLPA 146

Db    301 SDALTKV----RAAQTKINEAKA-------LLQNKEDNSQLVTSKNNLQSSVNQ------ 343

Qy    147 PLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQ 206

Db    344 -----VPSTTGMTQQSIDNYNAKKREAET----------EITAAQRVIDNGDATAQQ 385

Qy    207 VLDTGVAVQTYSARN---------AVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGF 257

Db    386 ISDENTTAQRNDLTNQISQATNLAAVESVKQSA-NSLDGAMGNLQTAINDKSGTLASQNF 444

Qy    258 GNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRM 317

Db    445 -------LDADEQKRNAYNQAVSNAE--TILNKQTGPNTAKTAVEQALNN---VNSAKHA 492

Qy    318 AGLPLDMATDAMGAVRSLVSASSLTQNG----LALAGGFAGVGKLQEMATKNITDPAT-- 371

Db    493 LNGTQNLNNAKQAAITAINGASDLNQKQKDALKAQANGAQRVSNAQDV-QRNATELNTAM 551

Qy    372 ---KAAVSQLTN-LAGSAAVFAGWTTA-ALTTDPAVKKAESFIQ--DTVKSTASSTTGYV 424

Db    552 GTLKHAIADKTNTLASSKYVNADSTKQNAYTT--KVTNAEHIISGTPTVVTTPSEVTAAA 609

Qy    425 ----------ADQTVKLAK----------TVKDMGGEA--------ITHT 446

Db    610 NQVNSAKQELNGDERLRVAKQNANTAIDALTQLNTPQKAKLKEQVGQANRLEDIQTVQTN 669

Qy    447 GASLRNTVNNLRQRPAREADIE--EGGTAASPS 477

Db    670 GQALNNAMKGLRDSIANETTVKASQNYTDASPN 702
```

RESULT 7
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12610
;   LENGTH: 5795
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12610

  Query Match             5.5%;  Score 131.5;  DB 10;  Length 5795;
  Best Local Similarity   21.6%;  Pred. No. 0.42;
  Matches  124;  Conservative   70;  Mismatches  226;  Indels  153;  Gaps   26;

Qy       3 INRRVQ-QPPVTATDSFRTASDASLASSSVRSV------------------SSDQQREI 42
           ||  ||      |||    ||   :||  |  ::                 :   :||:
Db    2462 INNAVQVANKRTATIKNGTAMPTNLAGGSTTTIPVTVTYNDGSTEEVQESIFTKADKREL 2521

Qy      43 NAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQ---ITALIETRASRLHFEGETPATI 99
             ::| | |    | |   :    |   :||||    ||  | :   :: |  ||  :
Db    2522 ITAKNHLDDPVSTDGKKPGTITQYNNA---IHNAQQQINTAKTEAQ-QVINNERATPQQV 2577

Qy     100 ADTFAKAEKLDRLATTTSGALRATPFAMASLLQ------------YMQPAINKGDWLPA 146
           :|   |     | | |     :|       |||              :| ::|:
Db    2578 SDALTKV----RAAQTKINEAKA-------LLQNKEDNSQLVTSKNNLQSSVNQ------ 2620

Qy     147 PLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQ 206
             |  :|    ::|    |  :  |              |  |  :  :  |:|
Db    2621 -----VPSTTGMTQQSIDNYNAKKREAET------------EITAAQRVIDNGDATAQQ 2662

Qy     207 VLDTGVAVQTYSARN---------AVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGF 257
           : |     |         || :|   |  |     | :   ::    |  |:  |
Db    2663 ISDENTTAQRNDLTNQISQATNLAAVESVKQSA-NSLDGAMGNLQTAINDKSGTLASQNF 2721

Qy     258 GNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRM 317
                 | |:  |    : : |    |:::     |  |:: |  :   :|: |
Db    2722 -------LDADEQKRNAYNQAVSNAE--TILNKQTGPNTAKTAVEQALNN---VNSAKHA 2769

Qy     318 AGLPLDMATDAMGAVRSLVSASSLTQNG----LALAGGFAGVGKLQEMATKNITDPAT-- 371
             ::      |: ::  || | |       | | |   |  |::  :| |:  |
Db    2770 LNGTQNLNNAKQAAITAINGASDLNQKQKDALKAQANGAQRVSNAQDV-QRNATELNTAM 2828

Qy     372 ---KAAVSQLTN-LAGSAAVFAGWTTA-ALTTDPAVKKAESFIQ--DTVKSTASSTTGYV 424
             | |::  || || |  | |   | ||   |  || |   |    || :| |  |
Db    2829 GTLKHAIADKTNTLASSKYVNADSTKQNAYTT--KVTNAEHIISGTPTVVTTPSEVTAAA 2886

Qy     425 -----------ADQTVKLAK------------------TVKDMGGEA--------ITHT 446
                    |: :::||                     :|:  |:|        :
Db    2887 NQVNSAKQELNGDERLRVAKQNANTAIDALTQLNTPQKAKLKEQVGQANRLEDIQTVQTN 2946

Qy     447 GASLRNTVNNLRQRPAREADIE--EGGTAASPS 477
           | :| | :  ||   | |  ::  :  | |||:
Db    2947 GQALNNAMKGLRDSIANETTVKASQNYTDASPN 2979


RESULT 8
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12713
;   LENGTH: 2344
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12713

  Query Match             5.4%;  Score 129.5;  DB 10;  Length 2344;
  Best Local Similarity   17.6%;  Pred. No. 0.17;
  Matches   86;  Conservative  107;  Mismatches  204;  Indels   91;  Gaps   18;

Qy       8 QQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADG 67
             ||    :|    : | ::  | |: : :|    :  :::  |:| | |:  |  ::|
Db     764 QQSQSVSTSKADSQSASTSTSGSIMTSTSASTSKSTSVS--LSDSVSASKSLSTSESN-- 819

Qy      68 QAAVDVHNAQITALIETRASRLHFEGE-TPATIADTF----AKAEKLDRLATTTSGALRA 122
                | ::   |:| : :::       |  : :|     |    :  || : ::|:|| :||
Db     820 ----SVSSSTSTSLVNSQSVSSSMSGSVSKSTSLSDFISNSSSTEKSESVSTSTSDSLRT 875

Qy     123 TPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLS 182
           :                                |:  :| : ||:: :  :: :  :| |    |
Db     876 S------------------------TSLSDSVSMSTSGSLSK--SQSLSTSTSD----S 904

Qy     183 ASPDR-LHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGAV 241
           ||  : : |: : |:      | :  : ::|   :: |  |::   :: : :   :   ::
Db     905 ASTSQSVSDSTSNSI-----STSESLSESGSTSESISISNSISNSVSASTSKLESQSTSI 959

Qy     242 DLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAI 301
           |  | :  :: :   :   | : |   | | |       :: :  :| :   : |
Db     960 SLSTSDSKSMSTSESLSD---STSTSDSVSGSLSVAG-----SQSVSTSTSDSMSTSEMI 1011

Qy     302 K-SASYSGAALNAGKRMAGLPLDMATDAMG----AVRSLVSASSLTQNGLALAGGFAGVG 356
             | | ||:   :  :   :  |:|   |    ::   :| |       |:|:   :|
Db    1012 SDSMSTSGSLAASDSKSMSVSSSMSTSQSGSTSESLSDSISTSDSDSKSLSLSTSQSG-- 1069

Qy     357 KLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKST 416
                :|   |  ::   :|:  : :|| :     :|:  |          ||   |  |
Db    1070 -----STSTSTSTSSSVRMSESQSTSGSMSTSQSDSTSIST---------SFSDSTSDSK 1115

Qy     417 ASSTTGYVADQTVKLAKTVK-DMGGEAITHTGASLRN------TVNNLRQRPAREADIEE 469
           ::||       :  ::::|     |  | | |  |  |      ::::       |:|
Db    1116 SASTA-----SSESISQSVSTSTSGSVSTSTSLSTSNSERTSTSMSDSTSLSTSESDSTS 1170

Qy     470 GGTAASPS 477
             |: | |
Db    1171 DSTSTSDS 1178


RESULT 9
US-09-815-242-5835
; Sequence 5835, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
```

```
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5835
;   LENGTH: 2434
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-5835

  Query Match             5.3%;  Score 129;  DB 10;  Length 2434;
  Best Local Similarity   19.7%;  Pred. No. 0.2;
  Matches  112;  Conservative   89;  Mismatches  237;  Indels  130;  Gaps   23;

Qy       3 INRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPA 62
           :|:: :    |  :: :  |||:    :   ::|       ||||   :   |:     |
Db      92 LNQKQKDALKTQVNNAQRVSDANNVQHTATELNSAMTALKAAIAD--KERTKASGNYVNA 149

Qy      63 DSADGQA----------AVDVHNAQIT------ALIETRASRLHFEGETPATIADTFAK 105
           |    ||                ||:|      |  :  |::    |:    :|   |
Db     150 DQEKRQAYDSKVTNAENIISGTPNATLTVNDVNSAASQVNAAKTALNGDNNLRVAKEHAN 209

Qy     106 AEKLDRLATTTSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQ 165
                 |  |  :       | | : :| |          :|  :  :: |: |  |
Db     210 --------NTIDGLAQLNNAQKAKLKEQVQSATTLDG--VQTVKNSSQTLNTAMKGLRDS 259

Qy     166 VGTKMMDRATGDLHYLSASPDRLHD------AMAASVKRHS-----PSLARQVLD----- 209
           :  :   :|    :|  |||:  ::       |  | : : |      |:   ||
Db     260 IANEATIKA--GQNYTDASPNNRNEYDSAVTAAKAIINQTSNPTMEPNTITQVTSQVTTK 317

Qy     210 -------TGVAVQTYSARNAVRTVLAPALASRPAVQGAVDLGVSMAG---GLAANAGFGN 259
                  :|    :|:| :  : :   | : |:  ::|   ::||      |       |
Db     318 EQALNGARNLAQAKTTAKNNLNNLTSINNAQKDALTRSIDGATTVAGVNQETAKATELNN 377

Qy     260 RLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGK---- 315
           : |:|:           |: |   : |  :   :|:|  : |||     |:|| |
Db     378 AMHSLQN----------GIND---ETQTKQTQKYLDAEPSKKSA--YDQAVNAAKAILT 421

Qy     316 RMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAA- 374
           : :|  :| |  |:       |:::    || |         | | : |    : | : |
Db     422 KASGQNVDKA--AVEQALQNVNSTKTALNGDAKLNEAKAAAK-QTLGTLTHINNAQRTAL 478

Qy     375 ---VSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASSTTGY-------- 423
              ::| ||: |   | |      |   | |: : |: |:|  | |   :  |
Db     479 DNEITQATNVEGVNTVKA----KAQQLDGAMGQLETSIRD--KDTTLQSQNYQDADDAKR 532

Qy     424 -VADQTVKLAKTV--KDMGG-----------EAITHTGASLRNTVN-------------- 455
            | |  | |:  |  ||           :|:|    :|    |
Db     533 TAYSQAVNAAATILNKTAGGNTPKADVERAMQAVTQANTALNGIQNLDRAKQAANTAITN 592

Qy     456 ----NLRQRPAREADIEEGGTAASPSEI 479
                | :|: | :|  :   |  :: : :
Db     593 ASDLNTKQKEALKAQVTSAGRVSAANGV 620


RESULT 10
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12996
;   LENGTH: 6281
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12996

  Query Match             5.3%;  Score 129;  DB 10;  Length 6281;
  Best Local Similarity   19.7%;  Pred. No. 0.72;
  Matches  112;  Conservative   89;  Mismatches  237;  Indels  130;  Gaps   23;

Qy       3 INRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPA 62
           :|:: :    |  :: :  |||:    :   ::|       ||||   :   |:     |
Db     871 LNQKQKDALKTQVNNAQRVSDANNVQHTATELNSAMTALKAAIAD--KERTKASGNYVNA 928

Qy      63 DSADGQA----------AVDVHNAQIT------ALIETRASRLHFEGETPATIADTFAK 105
           |    ||                ||:|      |  :  |::    |:    :|   |
Db     929 DQEKRQAYDSKVTNAENIISGTPNATLTVNDVNSAASQVNAAKTALNGDNNLRVAKEHAN 988

Qy     106 AEKLDRLATTTSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQ 165
                 |  |  :       | | : :| |          :|  :  :: |: |  |
Db     989 --------NTIDGLAQLNNAQKAKLKEQVQSATTLDG--VQTVKNSSQTLNTAMKGLRDS 1038

Qy     166 VGTKMMDRATGDLHYLSASPDRLHD------AMAASVKRHS-----PSLARQVLD----- 209
           :  :   :|    :|  |||:  ::       |  | : : |      |:   ||
Db    1039 IANEATIKA--GQNYTDASPNNRNEYDSAVTAAKAIINQTSNPTMEPNTITQVTSQVTTK 1096

Qy     210 -------TGVAVQTYSARNAVRTVLAPALASRPAVQGAVDLGVSMAG---GLAANAGFGN 259
                  :|    :|:| :  : :   | : |:  ::|   ::||      |       |
Db    1097 EQALNGARNLAQAKTTAKNNLNNLTSINNAQKDALTRSIDGATTVAGVNQETAKATELNN 1156

Qy     260 RLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGK---- 315
           : |:|:           |: |   : |  :   :|:|  : |||     |:|| |
Db    1157 AMHSLQN----------GIND---ETQTKQTQKYLDAEPSKKSA--YDQAVNAAKAILT 1200
```

```
Qy     316 RMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAA- 374
Db    1201 KASGQNVDKA--AVEQALQNVNSTKTALNGDAKLNEAKAAAK-QTLGTLTHINNAQRTAL 1257

Qy     375 ---VSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASSTTGY-------- 423
Db    1258 DNEITQATNVEGVNTVKA----KAQQLDGAMGQLETSIRD--KDTTLQSQNYQDADDAKR 1311

Qy     424 -VADQTVKLAKTV--KDMGG-----------EAITHTGASLRNTVN-------------- 455
Db    1312 TAYSQAVNAAATILNKTAGGNTPKADVERAMQAVTQANTALNGIQNLDRAKQAANTAITN 1371

Qy     456 ----NLRQRPAREADIEEGGTAASPSEI 479
Db    1372 ASDLNTKQKEALKAQVTSAGRVSAANGV 1399
```

RESULT 11
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Protiens of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 5.3%; Score 127.5; DB 10; Length 596;
Best Local Similarity 20.8%; Pred. No. 0.036;
Matches 115; Conservative 70; Mismatches 179; Indels 189; Gaps 25;

```
Qy      20 TASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQA---------- 69
Db      88 TAGQAELTAAQVRV----------AAAAYETAYGLT---VPPPVIAENRAELMILIATNL 134

Qy      70 ------AVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL--- 120
Db     135 LGQNTPAIAVNEAEYGEMWAQDAAAMFGYAAATATATATLLPFEEAPEM--TSAGGLLEQ 192

Qy     121 -----RATPFAMAS-LLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVG------T 168
Db     193 AAAVEEASDTAAANQLMNNVPQALQQ---LAQPTQGTTP--SSKLGGLWKTVSPHRSPIS 247

Qy     169 KMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLA 228
Db     248 NMVSMANNHMS-MTNSGVSMTNTLSSMLKGFAPAAAAQ------AVQT-AAQNGVR---- 295
```

```
Qy     229 PALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQL 288
Db     296 -AMSSLGSSLGSSGLG----GGVAANLG---RAASVGS----------------------L 326

Qy     289 SEENDWLEAYKAIKSAS--------YSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASS 340
Db     327 SVPQAWAAANQAVTPAARALPLTSLTSAAERGPGQMLGGLPVG----------------- 369

Qy     341 LTQNGLALAGGFAGVGKL--------QEMATKNITDPA---------------TKAAVSQ 377
Db     370 --QMGARAGGGLSGVLRVPPRPYVMPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQ 427

Qy     378 L--------TNLAGSAAVFAGWTTAALTTDP------------AVKKAESFIQDTVKSTA 417
Db     428 VGPQVVNINTKLGYNNAVGAG---TGIVIDPNGVVLTNNHVIAGATDINAFSVGSGQTYG 484

Qy     418 SSTTGYVADQTVKLAK--------TVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEE 469
Db     485 VDVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMGNS---------------- 528

Qy     470 GGTAASPSEIPFR 482
Db     529 GGQGGTPRAVPGR 541
```

RESULT 12
US-09-813-214A-9
; Sequence 9, Application US/09813214A
; Patent No. US20020177200A1
; GENERAL INFORMATION:
; APPLICANT: Tucker, Kenneth
; APPLICANT: Plosila, Laura
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE, GE
; TITLE OF INVENTION: SEQUENCE AND USES THEREOF
; FILE REFERENCE: 7969-089-999
; CURRENT APPLICATION NUMBER: US/09/813,214A
; CURRENT FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 08/968,685
; PRIOR FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 2122
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-813-214A-9

Query Match 5.3%; Score 127; DB 9; Length 2122;
Best Local Similarity 21.1%; Pred. No. 0.23;
Matches 116; Conservative 84; Mismatches 218; Indels 132; Gaps 27;

```
Qy      23 DASLASSSVRSVSSDQQREIN-----AIADYLTDHVFAAHKLPPADS--------ADGQA 69
Db     281 DNDVNAANVRAYAPDDNQPIDNRYKATFKNGATD-VFSIGNSNGNDSIRRKIINVGAGSA 339

Qy      70 AVDVHN-AQITALIETRASRLHFEGETPATIADTFAKAEK-LDRLATTTSGALRATPFAM 127
Db     340 DTDAVNVAQLKEAVRLANRQITFKGD------DSNNRVEKGLGKTLTITGGA------QT 387

Qy     128 ASLLQYMQPAINKGDWLPAPL-KPLTPL--------------ISGALSGAMDQVG----- 167
Db     388 SALTDHNIGVVQNGDGLKVQLAETLTSLKMVTTENLTANEKVTVGKTRLTTDKIGFTNDM 447

Qy     168 ------TKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLA--RQVLDTG-VAVQTYS 218
Db     448 NGIDESKPYLDKDTG----IHAGGQKITKLTAGVVDDDAATYGQLKKVNQTAESALQTFT 503

Qy     219 ARNAVRT--------VLAPALASRP--------------------AVQGAVDLGVSMAGG 250
Db     504 VKKVDKNGNDANDSKIITVGKNNKPDGTQVNTLKLKGENGVDVTTETNGTVTFGLNQNNG 563

Qy     251 L-AANAGFGNRLLSVQSRDHQR------GGALVLGLKDKEPKAQLSEENDWLEAYKAIKS 303
```

```
Db    564 LTVGNSTLNNDGLSVKNTNSNKQIQVGADGITFTDISNSKPGAGI--ENTTRITRDGIGF 621

Qy    304 ASYSGAALNAGK-RMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKL---Q 359

Db    622 ANNTG-SLDANKPRLTPTGINAGGKELTNVQSAINPAT---NG----------GQLDFMN 667

Qy    360 EMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASS 419

Db    668 RLSTANTEKSGSAATIKDLYNLSQVPLTFAGDTGPNVTK----KLGEILKVKGGKTTADD 723

Qy    420 TT----GYVADQ-----TVKLAKTVKDMGGEAITHTGASLRNTV---NNLRQRPAREADI 467

Db    724 LTKNNIGVVADSTDNSLTVKLAKTLSDLDAVNTKTLTASDKVTVDSGNNTAKLQNGDLTF 783

Qy    468 EEGGTAASPS 477

Db    784 SKQNTGATPA 793
```

RESULT 13
US-09-379-931-7
; Sequence 7, Application US/09379931
; Patent No. US20020009792A1
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. US20020009792A1ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/09/379,931
; CURRENT FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614,377
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: US 08/194,290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895,367
; PRIOR FILING DATE: 1992-06-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1026
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-379-931-7

Query Match 5.2%; Score 126.5; DB 10; Length 1026;
Best Local Similarity 22.2%; Pred. No. 0.092;
Matches 113; Conservative 56; Mismatches 177; Indels 163; Gaps 23;

```
Qy     12 VTATDSFRTASDASLASSSVRSVSSDQQREINAIADY---------------------- 48

Db    206 VSGIGGYATATAAMINDLSDGALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTA 265

Qy     49 -----LTDHVFAAHKLPPADSADGQAAVDVHN----AQITAL--------IETRASRLHF 91

Db    266 NNDTFVAGEVAGAATLTVGDTLSGGAGTDVLNWVQAAAVTALPTGVTISGIET----MNV 321

Qy     92 EGETPATIADTFAKAEKLDRLATTTSGA----------LRATPFAMASLLQYMQPAINK 140

Db    322 TSGAAITL-NTSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGRAN- 379

Qy    141 GDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHS 200

Db    380 ----------VTVASTGVTSGT-TTVGAN--SAASG-------------TVSVSVANSS 412

Qy    201 PSLARQVLDTGVAVQT--YSARNAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFG 258

Db    413 TTTTGAIAVTGGTAVTVAQTAGNAVNTTLTQADVTVTGNSSTTAVTVTQTAAATAGATVA 472

Qy    259 NRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMA 318

Db    473 GRV----------NGAVTI------------------------TDSAAASATTAGK--- 494
```

```
Qy    319 GLPLDMATDAMGAV-RSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQ 377

Db    495 -----IATVTLGSFGAATIDSSALTTVNLSGTGTSLGIGR------GALT--ATPTANTL 541

Qy    378 LTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQD---TVK---STASSTTGYVADQTVKL 431

Db    542 TLNVNG-----------LTTTGAITDSEAAADDGFTTINIAGSTASST---IASLVAAD 586

Qy    432 AKTVKDMGGEAIT---HTGASLRN-TVNN 456

Db    587 ATTLNISGDARVTITSHTAAALTGITVTN 615
```

RESULT 14
US-09-841-786-4
; Sequence 4, Application US/09841786
; Patent No. US20020054883A1
; GENERAL INFORMATION:
; APPLICANT: NAGARAJA, T. G.
; APPLICANT: STEWART, GEORGE C.
; APPLICANT: NARAYANAN, SANJEEV K.
; APPLICANT: CHENGAPPA, M. M.
; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN
; TITLE OF INVENTION: VACCINE AND PREPARATION THEREOF
; FILE REFERENCE: 30296
; CURRENT APPLICATION NUMBER: US/09/841,786
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 09/558,257
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Fusobacterium necrophorum
US-09-841-786-4

Query Match 5.2%; Score 125; DB 10; Length 714;
Best Local Similarity 20.3%; Pred. No. 0.072;
Matches 113; Conservative 91; Mismatches 220; Indels 132; Gaps 25;

```
Qy     14 ATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQ----- 68

Db     73 AVNNIQSKTSALVKNSDIRNAN---KFKVNALSG--GTQVAAGAGLEAVKESGGQGKSYL 127

Qy     69 ----AAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLAT--------TT 116

Db    128 LGTSASINLVNNEVSAKSENNT----VAGESESQKMDVDVTAYQADTQVTGALNLQAGKS 183

Qy    117 SGALRATPFAMASLLQYMQPAINKGDWL---PAPLKPLTPLISGALSGAMDQVGTKMMDR 173

Db    184 NGTVGAT-VTVAKLNNKVNASISGGRYTNVNRADAKAL--LATTQVTAAVTTGGTISSGA 240

Qy    174 ATGDLHYLSASPDRLHDAMAASVKRHSPS-------LARQVLDTGVAVQTYSARNAVRTV 226

Db    241 GLGN-YQGAVSVNKIDNDVEASVDKSSIEGANEINVIAKDVKGSSDLAKEYQ-------- 291

Qy    227 LAPALASRPAVQGAVDLGVSMAGG------------LAANAGFGNRLLSVQSRDHQRGG- 273

Db    292 ---ALLNGKDKKYLEDRGINTTGNGYYTKEQLEKAKKKEGAVIVNAALSVAGTDKSAGGV 348

Qy    274 ALVLGLKDKEPKAQLSEEN-----DWLEAYKAIKSASYSGAALNAGKRMA---------- 318

Db    349 AIAVNTVKNKFKAELSGSNKEAGEDKIHAKHVNVEAKSSTVVVNAASGLAISKDAFSGMG 408

Qy    319 -GLPLDMATDAMGAV-RSLVSASSLTQN------GLALAGGFAG-----VG--------- 356

Db    409 SGAWQDLSNDTIAKVDKGRISADSLNVNANNSILGVNVAGTIAGSLSTAVGAAFANNTLH 468

Qy    357 -KLQEMATKNITDP----ATKAAV-----SQLTNLA--GSAAVFAGWTTAALTTDPAVKK 404

Db    469 NKTSALITGTKVNPFSGKNTKVNVQALNDSHITNVSAGGAASIKQAGIGGMVSVNRGSDE 528
```

```
Qy    405 AESFIQDTVKSTASSTTGYVADQTVKLAKTVKDMGGEA----------ITHTGASLRNT 453
          |: : |:     ||       ||      ||:  : | |           : ||    :: :
Db    529 TEALVSDSEFEGVSSFNVDAKDQ-----KTINTIAGNANGGKAAGVGATVAHTNIGKQSV 583

Qy    454 ---VNNLRQRPAREAD 466
              | | :   | : |
Db    584 IAIVKNSKITTANDQD 599


RESULT 15
US-09-841-786-1
; Sequence 1, Application US/09841786
; Patent No. US20020054883A1
; GENERAL INFORMATION:
;  APPLICANT: NAGARAJA, T. G.
;  APPLICANT:  STEWART, GEORGE C.
;  APPLICANT:  NARAYANAN, SANJEEV K.
;  APPLICANT:  CHENGAPPA, M. M.
;  TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN
;  TITLE OF INVENTION:  VACCINE AND PREPARATION THEREOF
;  FILE REFERENCE: 30296
;  CURRENT APPLICATION NUMBER: US/09/841,786
;  CURRENT FILING DATE:  2001-04-24
;  PRIOR APPLICATION NUMBER: 09/558,257
;  PRIOR FILING DATE: 2000-04-25
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 3241
;   TYPE: PRT
;   ORGANISM: Fusobacterium necrophorum
US-09-841-786-1

  Query Match             5.2%;  Score 125;  DB 10;  Length 3241;
  Best Local Similarity   20.3%;  Pred. No. 0.58;
  Matches  113;  Conservative   91;  Mismatches  220;  Indels  132;  Gaps   25;

Qy     14 ATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQ----- 68
          | :: :: : | : :| :|: :   : ::||::      | |   |      : ||
Db   1257 AVNNIQSKTSALVKNSDIRNAN---KFKVNALSG--GTQVAAGAGLEAVKESGGQGKSYL 1311

Qy     69 ----AAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLAT--------TT 116
              |:::: | :::|  |         ||: :   |    | : |   |         :
Db   1312 LGTSASINLVNNEVSAKSENNT----VAGESESQKMDVDVTAYQADTQVTGALNLQAGKS 1367

Qy    117 SGALRATPFAMASLLQYMQPAINKGDWL---PAPLKPLTPLISGALSGAMDQVGTKMMDR 173
          :| : ||   :| |   :  :|: | :      |  | |  | :  :: |:   ||
Db   1368 NGTVGAT-VTVAKLNNKVNASISGGRYTNVNRADAKAL--LATTQVTAAVTTGGTISSGA 1424

Qy    174 ATGDLHYLSASPDRLHDAMAASVKRHSPS-------LARQVLDTGVAVQTYSARNAVRTV 226
            |: :  : | ::: : : ||| : |          :|: |  :    : |
Db   1425 GLGN-YQGAVSVNKIDNDVEASVDKSSIEGANEINVIAKDVKGSSDLAKEYQ-------- 1475

Qy    227 LAPALASRPAVQGAVDLGVSMAGG-----------LAANAGFGNRLLSVQSRDHQRGG- 273
            || :    :   | |::  |                |   |  |||   |   ||
Db   1476 ---ALLNGKDKKYLEDRGINTTGNGYYTKEQLEKAKKKEGAVIVNAALSVAGTDKSAGGV 1532

Qy    274 ALVLGLKDKEPKAQLSEEN-----DWLEAYKAIKSASYSGAALNAGKRMA---------- 318
          |: :     : ||:||  |      | : |      |  |   :||   :|
Db   1533 AIAVNTVKNKFKAELSGSNKEAGEDKIHAKHVNVEAKSSTVVVNAASGLAISKDAFSGMG 1592

Qy    319 -GLPLDMATDAMGAV-RSLVSASSLTQN------GLALAGGFAG-----VG--------- 356
           |   |:: | :  | :  :|| ||  |       |: :||  ||      ||
Db   1593 SGAWQDLSNDTIAKVDKGRISADSLNVNANNSILGVNVAGTIAGSLSTAVGAAFANNTLH 1652

Qy    357 -KLQEMATKNITDP----ATKAAV-----SQLTNLA--GSAAVFAGWTTAALTTDPAVKK 404
           |   : |    :|      ||  |      | :||::  |:|::         :: :    :
Db   1653 NKTSALITGTKVNPFSGKNTKVNVQALNDSHITNVSAGGAASIKQAGIGGMVSVNRGSDE 1712

Qy    405 AESFIQDTVKSTASSTTGYVADQTVKLAKTVKDMGGEA----------ITHTGASLRNT 453
          |: : |:     ||       ||      ||:  : | |           : ||    ::
Db   1713 TEALVSDSEFEGVSSFNVDAKDQ-----KTINTIAGNANGGKAAGVGATVAHTNIGKQSV 1767

Qy    454 ---VNNLRQRPAREAD 466
              | | :   | : |
Db   1768 IAIVKNSKITTANDQD 1783


Search completed: January  2, 2003, 15:22:06
Job time : 17.4913 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:16:50 ; Search time 29.9692 Seconds
(without alignments)
3341.397 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHINRRVQQPPVTATDSFRT.....IEEGGTAASPSEIPFRPMRS 486

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2412	100.0	486	2	Q9JP34	Q9jp34 pseudomonas
2	1653	68.5	484	2	Q8RP03	Q8rp03 pseudomonas
3	490	20.3	518	16	Q8XQF0	Q8xqf0 ralstonia s
4	153	6.3	1545	16	Q9RDQ1	Q9rdq1 streptomyce
5	151.5	6.3	516	16	Q8XSS2	Q8xss2 ralstonia s
6	150.5	6.2	2055	2	O85472	O85472 abiotrophia
7	149	6.2	1953	16	Q98HJ2	Q98hj2 rhizobium l
8	148	6.1	1203	5	Q9N5K0	Q9n5k0 caenorhabdi
9	147.5	6.1	1363	2	Q9F5K8	Q9f5k8 rhodobacter
10	147.5	6.1	1822	2	Q07290	Q07290 streptococc
11	147	6.1	1156	16	Q9Z5A4	Q9z5a4 streptomyce
12	147	6.1	6048	2	Q93H87	Q93h87 streptomyce
13	145.5	6.0	973	16	Q8XDQ4	Q8xdq4 escherichia
14	144.5	6.0	2124	16	Q98M03	Q98m03 rhizobium l
15	144.5	6.0	2232	5	P91365	P91365 caenorhabdi
16	144	6.0	515	16	Q8XT14	Q8xt14 ralstonia s
17	143.5	5.9	1713	3	Q8TGE1	Q8tge1 saccharomyc
18	142.5	5.9	949	4	Q9Y4G6	Q9y4g6 homo sapien
19	142	5.9	2016	5	Q9BIT0	Q9bit0 plectreurys
20	141.5	5.9	491	2	Q93GT1	Q93gt1 campylobact
21	141.5	5.9	3381	2	Q9KX33	Q9kx33 streptococc
22	139	5.8	1794	9	Q9T1A7	Q9t1a7 bacteriopha
23	138.5	5.7	2155	16	Q8UFP9	Q8ufp9 agrobacteri
24	138	5.7	506	2	Q53834	Q53834 salmonella
25	138	5.7	825	17	Q8TJH9	Q8tjh9 methanosarc
26	137.5	5.7	5636	5	Q9N9M2	Q9n9m2 leishmania
27	137	5.7	506	2	Q53836	Q53836 salmonella
28	136.5	5.7	491	2	Q8RTY4	Q8rty4 campylobact
29	136.5	5.7	653	2	Q9X6N1	Q9x6n1 rhizobium l
30	136	5.6	597	16	Q9S2K2	Q9s2k2 streptomyce
31	136	5.6	4776	16	Q97P71	Q97p71 streptococc
32	135.5	5.6	1731	16	Q8U8W4	Q8u8w4 agrobacteri
33	135.5	5.6	2479	16	Q9A988	Q9a988 caulobacter
34	135	5.6	718	16	Q9JZD0	Q9jzd0 neisseria m
35	135	5.6	1128	17	O52009	O52009 halobacteri
36	134	5.6	492	2	Q93GT3	Q93gt3 campylobact
37	134	5.6	2271	16	Q99QY4	Q99qy4 staphylococ
38	133.5	5.5	466	16	Q98LA8	Q98la8 rhizobium l
39	133.5	5.5	5644	2	Q93NX8	Q93nx8 streptomyce
40	133	5.5	506	2	Q53837	Q53837 salmonella
41	133	5.5	1208	16	Q9PFI9	Q9pfi9 xylella fas
42	132.5	5.5	487	2	Q93GT2	Q93gt2 campylobact
43	132.5	5.5	1336	16	Q9ACV2	Q9acv2 streptomyce
44	132	5.5	540	9	Q9G0H8	Q9g0h8 roseophage
45	132	5.5	1762	2	O52546	O52546 amycolatops

ALIGNMENTS

RESULT 1
Q9JP34
ID Q9JP34 PRELIMINARY; PRT; 486 AA.
AC Q9JP34;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 50.7 kDa protein.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=98422476; PubMed=9748456;
RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA Collmer A.;
RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
RT to harpins and pectate lyases and can elicit the plant hypersensitive
RT response and bind to pectate.";
RL J. Bacteriol. 180:5211-5217(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=20243785; PubMed=10781092;
RA Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
RA Petnicki-Ocwieja T., van Dijk K., Collmer A.;
RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT mosaic structure composed of a cluster of type III secretion genes
RT bounded by exchangeable effector and conserved effector loci that
RT contribute to parasitic fitness and pathogenicity in plants.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Ramos A.R., Rehm A.H., Collmer A.R.;
RT "Pseudomonas syringae pv. tomato DC3000 hrpL through hrcU.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF232004; AAF71504.1; -.
KW Hypothetical protein.
SQ SEQUENCE 486 AA; 50709 MW; 91CFBC03397F8AD1 CRC64;

Query Match 100.0%; Score 2412; DB 2; Length 486;
Best Local Similarity 100.0%; Pred. No. 1.1e-122;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60
Db       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60

Qy      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120
Db      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120

Qy     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180
Db     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180

Qy     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240
Db     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240

Qy     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300
Db     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300

Qy     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360
Db     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360

Qy     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420
Db     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420

Qy     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEIP 480
Db     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEIP 480

Qy     481 FRPMRS 486
Db     481 FRPMRS 486
```

RESULT 2
Q8RP03
ID Q8RP03 PRELIMINARY; PRT; 484 AA.
AC Q8RP03;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Type III effector HopPtoA1Pma.
OS Pseudomonas syringae pv. maculicola.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=59511;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ES4326;
RX MEDLINE=21862332; PubMed=11872842;
RA Guttman D.S., Vinatzer B.A., Sarkar S.F., Ranall M.V., Kettler G.,
RA Greenberg J.T.;
RT "A Functional Screen for the Type III (Hrp) Secretome of the Plant
RT Pathogen Pseudomonas syringae.";
RL Science 295:1722-1726(2002).
DR EMBL; AF458051; AAL84253.1; -.
SQ SEQUENCE 484 AA; 50520 MW; 4DB012DD5688163E CRC64;

Query Match 68.5%; Score 1653; DB 2; Length 484;
Best Local Similarity 69.1%; Pred. No. 1.1e-81;
Matches 337; Conservative 50; Mismatches 93; Indels 8; Gaps 3;

```
Qy       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60
Db       1 MYINRSISSQSSIGTESFHSAQ--SVASSSVRVLSEAEQTKLNDITHYLTDHVFAAHRLP 58

Qy      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120
Db      59 VNDINVDQDVVHAHNVEIGKIISARALRLLDEGETALSIGETFAKAEKFDRMASAASGAL 118

Qy     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180
Db     119 RATPFAAASVLQYMQPAINKGDWLPTPLKPLAPLVSGALSGVMDQVGTGVMNRATGDLHY 178

Qy     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240
Db     179 LSTAPEKLHDAMAASVKRHKPGVMRQAADLGIAVQTYTARNALRTVLAPALASRPAVQGA 238

Qy     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300
Db     239 VDISVSAAGSLVANAGFSERMHTVQARDHLRGGAFVLGIKDKQPKADLSEETDWLDAYRA 298

Qy     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360
Db     299 IKSASYSGAALNAGKRVAGLPLDVATDGLKAVRSLVSATSLMQNGLVMAGGFAGVGKLQE 358

Qy     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420
Db     359 MATKNITHPATKAAVSQLTNLAGSAPVFSAWTTAAVATDPAVKKAESFLQDTVKTTVSNG 418

Qy     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAA---SPS 477
Db     419 TSALADKTVKLAKAGIDASAERIATTGASLRDT---LRRRTVREPDIEEGGVAAGTGSPV 475

Qy     478 EIPFRPMR 485
Db     476 AVPFEAMR 483
```

RESULT 3
Q8XQF0
ID Q8XQF0 PRELIMINARY; PRT; 518 AA.
AC Q8XQF0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative transmembrane protein.
GN RSP1277 OR RS05322.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646083; CAD18428.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 518 AA; 53581 MW; 82D486715F0F60EE CRC64;

Query Match 20.3%; Score 490; DB 16; Length 518;
Best Local Similarity 34.1%; Pred. No. 7.8e-19;
Matches 147; Conservative 58; Mismatches 158; Indels 68; Gaps 12;

```
Qy      17 SFRTASD--------ASLASSSVRSVSSDQQ--------------REIN-----AIADY 48
```

```
Db      98 SERSNSDVSQWFSAAASIRAQSIGSIESNSPPGGSPPHTADPGQTGRGIDIPHWMPIADQ 157

Qy      49 LTDHVFAAHKLPPADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEK 108
           | | :|  |         ||       ::||    ||: || ||    ||||||:|    ||
Db     158 LRD-IFRPH-------FDG-----TNDAQFEILIQQRAERLQAMGETPATVAAVLAKGAN 204

Qy     109 LDRLATTTSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGT 168
           ||||  ||  | :|: || :|| |  ::  |:          |             ||  ||   |  |
Db     205 RDRLAQTTVGFVRSVPFGIASRLFDVKQALTAF----AKTTARVGATVGAGSGTADAFGG 260

Qy     169 KMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLA 228
            :: :|| :  :|:|||| |   || : |   ||| |   :  :| |||| || :|| :|
Db     261 TLLGKATSNTQWLAASPDHLEPVMAQAHKAVQPSLGRLAAEVSLAFQTYSLRNVIRTGVA 320

Qy     229 P----ALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEP 284
           |    || :  |    ||  ::  ||  | |     :  :    |: |   :||
Db     321 PLATHALGAARAAN--VDSWIAAVGGPVAGAAAYMAMQHMNETHHRTGAEYLLG------ 372

Qy     285 KAQLSEENDWLEAYKAIKSASYSGAALNAGKRMAGLPL-DMATDAMGAVRSLVSASSLTQ 343
                 || : :| :| ::::   : |||| | | : |:||: : | ||| :|::| :
Db     373 ------RTDWEDQFKQLKQSTWTDPLVGAGKRTAKLLVADLATETLAATRSLFTATNLIK 426

Qy     344 NGLALAGGFAGVGKLQEMATKNITD----PATKAAVSQLTNLAGSAAVFAGWTTAALTTD 399
           |  |||||||||    |  | |  |:     |  ||| :  :   || |:| |||| :
Db     427 NMGALAGGFAGVLTAQTAAGKAATEAGYTEAAVAAVRRAVSTVLSAPVYAAWTTADVMAG 486

Qy     400 PAVKKAESFIQ 410
           ||:  |   ||
Db     487 PAIDAAAGHIQ 497
```

RESULT 4
Q9RDQ1
ID Q9RDQ1 PRELIMINARY; PRT; 1545 AA.
AC Q9RDQ1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative secreted protein.
GN SCO2383 OR SC4A7.11.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL133423; CAB62715.1; -.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00652; Ricin_B_lectin; 3.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
SQ SEQUENCE 1545 AA; 161520 MW; 81EF325143593AEA CRC64;

Query Match 6.3%; Score 153; DB 16; Length 1545;
Best Local Similarity 22.6%; Pred. No. 5.5;
Matches 113; Conservative 66; Mismatches 202; Indels 118; Gaps 20;

```
Qy'      6 RVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAI-----------ADYLTDHV 53
           | :    |||:  : |  |   : ::|  : | ::: ||              |    |:
Db     745 RAETADTTATEKRKGAEAARDKAKALRDDAWDAEQKANAARAKADAKEAYAQASASEDNA 804

Qy      54 FAAHKLPPADSADGQAAVDVHNAQITALIET-RASRLHFEGETPATIADTFAKAEKLDRL 112
             | |   |: ||  || |   | : |  |  :|::   : :  || |:  ||  : |
Db     805 QEARK--AANDADA-AADDAETAAVAARSEADKATKAAADADAAATRAEAAAKRSRADSD 861

Qy     113 AT-----TTSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVG 167
           |       |   |:|    | |  ::  : |                          :
Db     862 AAQAAKLTADAAVRTATSAAADAIKASKTAATAA----------------------RTA 898

Qy     168 TKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQT----YSARNAV 223
            :: | |  : |    : |   :| ||  :  : :||      || |  |      | |:
Db     899 VELADDA--EQH----AADAKKEADAAKAEAVT-ALAGANESTGYAYTTAQAAVDAGNSA 951

Qy     224 RTVLAPALASRPAVQ-GAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDK 282
             | |||      |:| |:  :    | |||   |  :: :: |    |:  |
Db     952 AQVAAPA---NDAIQLGSPYVTTDSAAGLAVLTGQSSKTIAEQ----QQAVA-------- 996

Qy     283 EPKAQLSEEN------------DWLEAYK-AIKSASYSGAALNAGKRMAGLPLDMATDA 328
           |  || :||:             |   ||  | ::  |:  | |: |   |   : |: |
Db     997 EAHAQNAEESAAQAQSVANAASGDSKAAYTLAAEALGYAADARNSAKEALGYSAEAASYA 1056

Qy     329 MGAVRSLV----------------SASSLTQNGLALAGGFAGVGKLQEMATKNITDPAT 371
             | :||                 ||:   :     |   |     |  | ::  | |
Db    1057 TQAAQSLTRTIAYDTQATKDAAAADSAAGRAEGHAEDARDSADAAALDAEAARSAADTAE 1116

Qy     372 KAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASSTTGYVADQTVKL 431
           :||       |  ||  |   | |   : | | |: : :  ::         |:|  |
Db    1117 QAAED-----ARDAADHA--ATEAAAAEEAAKDAQKYAESAQQA---------AEQAEKE 1160

Qy     432 AKTVKDMGGEAITHTGASL 450
           |   :   |  :  ||| :
Db    1161 ANAEQIDKGTVVDQTGAPI 1179
```

RESULT 5
Q8XSS2
ID Q8XSS2 PRELIMINARY; PRT; 516 AA.
AC Q8XSS2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Probable flagellar hook-length control protein.
GN FLIK OR RSP0395 OR RS00823.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,

```
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646078; CAD17546.1; -.
DR   InterPro; IPR000104; Antifreeze_1.
DR   PRINTS; PR00308; ANTIFREEZEI.
KW   Plasmid; Complete proteome.
SQ   SEQUENCE   516 AA;  50024 MW;  8969D64D4F43ECF4 CRC64;

  Query Match             6.3%;  Score 151.5;  DB 16;  Length 516;
  Best Local Similarity   24.0%;  Pred. No. 1.5;
  Matches  121;  Conservative   46;  Mismatches  208;  Indels  129;  Gaps   19;

Qy      10 PPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQA 69
           |   |||      | |  |:   :: ||       | ::| |     || :   | ||  :|
Db       5 PNTAATDLGAMLSSAGTAADKSQAASSGADTFGNLLSDRLNQE--AARRQTDAQSAGSRA 62

Qy      70 AVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMAS 129
           |          | |              | | | |   : :|
Db      63 A---------------ARR--------------ADATKKDAAKASAAGN---------- 82

Qy     130 LLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLH 189
             |  : |    |  ||         : | |   |  :|           :|| |
Db      83 -DQDARQADRNADAAPA---------TAAASATQDAQASK-----------AASSDDTQ 120

Qy     190 DAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGAV--DLGVSM 247
           |  ||:     |:          :| |  :|| |::   ||    |:|  | |    |    :
Db     121 DKTAAAQAGVDPA-------AQLAAQIEAARQAMQQAAAP--VSQPGTQVATADDAAAQL 171

Qy     248 AGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYS 307
           | | |  |      |: |  :   | |       :| | :::       :| :: :
Db     172 AAGKAGAATTDAAALAALSAANAAAGKLT------DPAATVAQP-------QAAQADALK 218

Qy     308 GAALNAGKRMAGLPL-DMATDAMG------------AVRSLVSASSLTQ-----NGLAL 348
             ||    :      || | :| |                ||| | ||||  |       | |
Db     219 AAAGKEAQAAKPTPLPDASTVAQALARTRGTTDNTPPAVRHLGSASSSAQGAGSERGSAA 278

Qy     349 AGGFAGVGKLQEMATKNITDPATKAAVSQLTN----LAGSAAVFAGWTTAALTTDPAVKK 404
           ||| |   : ||    :    |  |    |      |: | | |  | ||     ||
Db     279 HRTEAGQGTQAQAATAATAETGRNADTSGNPNTQADAQGAGAKFDGVLTRA--NGDAVAT 336

Qy     405 AESFIQDTVKSTASSTTGYVADQTVKLAKTVKDMGGEAITHTGASL-------RNTVNNL 457
           | :|   ||    :: :   |  |   | |:   |  |  || ||        | :   |
Db     337 APTF---TVGGAGTAGSAASAAVTPAQAHTLPTFGDAAWPHTMASQLAYMQVHRQSSAEL 393

Qy     458 RQRPAR------EADIEEGGTAAS 475
           :  ||       : :|: |   ||
Db     394 QLSPAELGGLHVKLEIDNGAVNAS 417


RESULT 6
O85472
ID   O85472      PRELIMINARY;      PRT;  2055 AA.
AC   O85472;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Extracellular matrix binding protein (Fragment).
GN   EMB.
OS   Abiotrophia defectiva.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC   Aerococcaceae; Abiotrophia.
OX   NCBI_TaxID=46125;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NVS-47;
RX   MEDLINE=99081722; PubMed=9864195;
RA   Manganelli R., van de Rijn I.;
RT   "Characterization of emb, a gene encoding the major adhesin of
RT   Streptococcus defectivus.";
RL   Infect. Immun. 67:50-56(1999).
DR   EMBL; AF067776; AAD03320.1; -.
DR   TIGRFAMs; TIGR01168; YSIRK_signal; 1.
FT   NON_TER    2055   2055
SQ   SEQUENCE   2055 AA;  215642 MW;  9699C11DDE93E2FD CRC64;

  Query Match             6.2%;  Score 150.5;  DB 2;  Length 2055;
  Best Local Similarity   22.2%;  Pred. No. 11;
  Matches  126;  Conservative   73;  Mismatches  246;  Indels  123;  Gaps   21;

Qy       3 INRRVQQPPVTA------TDSFRTASDASLA----SSSVRSVSSDQQREINAIADYLTDH 52
           ||    |:| :|         |  ||| ||::|     ::: ::|:| : : :||: :  |
Db     726 INEISQRPDLTREEKQAFMDQVRTARDAAMAKVASAANNQAVTSARDQGLNAVNNLPT-- 783

Qy      53 VFAAHKLPPA-----DSADGQAAVDVHNAQITALIETRASRLHFEGETPATIA------- 100
             | | | |       :|| :       || :||  :  | |      :| |  |
Db     784 --PAAKYPEALGHVRQAADAKRQAIRDNANLTAEEQADALRQVDAAQTAAEAAINQNHTN 841

Qy     101 DTFAKAEKLD-RLATTTSGALRATPFAMASLLQY---MQPAINKGDWLPAPLK-PLTPLI 155
             | |||:    :      :    |: | |  :| |      : |||  : |    |       :
Db     842 ATLAKADSDGVKAINDINPQPRSKPAANQALEQVAAAKRQAINNNNQLTDEEKAQAIQQV 901

Qy     156 SGALSGAMDQV-------GTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVL 208
             ||: |  ||        |     :  |     : :|     | | :    :     | |
Db     902 DQALANAKTQVQAANDNNGVNQA-KTAGTTAINNINPQGTQKAQAIAAIEAAEQAKRLEL 960

Qy     209 DTGVAVQTYSARNAVRTVLAPALASRPAVQ---------GAVDLGVSMAGGL-------- 251
               : |     ||:  : | | |:: ||          || | ||:    |:
Db     961 QGRNDLTTEERNNALADLTAKAQAAKDAVNQARNNTGVAGAKDNGVAQIQGINPTAVVKP 1020

Qy     252 -AANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWL-EAYKAIKSASYSGA 309
            | ||      :   :|| :        | |:|  | : :  |    :|  ||  |  :|
Db    1021 DARNA------IDQAARDKEAEFQANTKLTDEEKAAAIKKVQDAARDAKAAIDRAGSNGD 1074

Qy     310 ALNA-------------------GKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALA 349
             ||                    |  |   :  | ||  |  :: : ::|||   |
Db    1075 VNNAVNQGKAAIQAIKALDDSQPSAKDTAKAAIQNAADAKKA--AITANNALTQEEKA-- 1130

Qy     350 GGFAGVGKLQEMATKNITDPATKAAVS----------------QLTNLAGSAAVFAGWTT 393
                   | | : | | ||  :                    |  ::   ||
Db    1131 -----------AAIKQVEDEAAKAQAAVDASRSKADVDRAKDQGLQKISDVPAVQPPKLN 1179

Qy     394 AALTTDPAVKKAESFIQDTVKST---ASSTTGYVADQTVKLAKTVKDMGGEA-ITHTGAS 449
           |    | |    :: | :    |      :    | ::  |  : : |    | :    |
Db    1180 AIAAVDQAATDKKAVINNDTTLTQEEKEAAIRKVDEEAAKARQAINDATSNADVAAKQAQ 1239

Qy     450 LRNTVNNLRQRP----AREADIEEGGTA 473
              :||: | |     | :| :|:   |
Db    1240 GTQAINNVPQTPAAKNAAKAAVEQAADA 1267


RESULT 7
Q98HJ2
ID   Q98HJ2      PRELIMINARY;      PRT;  1953 AA.
AC   Q98HJ2;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein mll2848.
GN   MLL2848.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
```

```
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
DR   EMBL; AP003000; BAB49874.1; -.
DR   InterPro; IPR000911; Ribosomal_L11.
DR   PROSITE; PS00359; RIBOSOMAL_L11; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   1953 AA;  184557 MW;  892BFA8B687B35E2 CRC64;

  Query Match             6.2%;  Score 149;  DB 16;  Length 1953;
  Best Local Similarity   23.1%;  Pred. No. 12;
  Matches  123;  Conservative   53;  Mismatches  216;  Indels  140;  Gaps   23;

Qy       2 HINRRVQQPPVTATDSF-------RTASDASLASSSVRSVSSDQQREINAIADYLTDHVF 54
Db     829 HANSTLADSSATGTDSVAIGPTSTATASSAIAAGSNANASGAN----ASAIGTSSVASAL 884

Qy      55 AAHKLPPADSADGQAAVDV-HNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLA 113
Db     885 DATAMGFLSKASGQFSTAVGANANATALSSTAIGQNALASGVQATALGKQANASASDALA 944

Qy     114 -------------TTTSGALRA----TPFAMASLLQYMQPAINKGDWL----PAPLKPLT 152
Db     945 LGANSTAGNAGDVALGSGSVTAVAVGTPNAVINGTTYAFQGINPASTVSIGAPGAERTLT 1004

Qy     153 PLISGALSG---------------AMDQVGTKMMDRAT-GDLHYLSASPDRLHDAMA-- 193
Db    1005 NLAAGRISGSSTDAVNGSQLFATNQAVDAIGTTVNNISTGGGIKYFHAN-STLADSSATG 1063

Qy     194 ---ASVKRHSPSLARQVLDTG----------VAVQTYSARNAVRTVL------------- 227
Db    1064 TDSVAIGPTSTATASSAIAAGSNANASGANASAIGTSSVASALDATAMGFLSKASGQFST 1123

Qy     228 -------APALASRPAVQGAVDLGV-SMAGGLAANAGFGNRL-LSVQSR-----DHQRGG 273
Db    1124 AVGANANATALSSTAIGQNALASGVQATALGKQANAGASDALALGANSTAGNAGDVALGS 1183

Qy     274 ALVLGLKDKEPKAQLSEENDWLEAYKAIKSAS-----YSGAALNAGKRMAGLPLDMATDA 328
Db    1184 GSVTAVAVGTPNAVI---NGTTYAFQGINPASTVSIGAPGAERTLTNLAAGRISGSSTDA 1240

Qy     329 MG-----AVRSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAG 383
Db    1241 VNGSQLFATNQAVDAIGTTVNN--LGGSVTNLGNV------------------VNNIAG 1279

Qy     384 SAAVFAGWTTA-------ALTTDPAVKKAESFIQDTVKSTASSTTGYVADQT 428
Db    1280 DTS--TAYTDANGIGIRYARTNEAGLAQTDSFAQ----GLGSTAVGYQASAT 1325


RESULT 8
Q9N5K0
ID   Q9N5K0      PRELIMINARY;      PRT;  1203 AA.
AC   Q9N5K0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical 120.1 kDa protein.
GN   H43E16.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Miller N., Carter T.;
RT   "The sequence of C. elegans cosmid H43E16.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006669; AAF39909.1; -.
DR   InterPro; IPR000082; SEA_domain.
DR   SMART; SM00200; SEA; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   1203 AA;  120104 MW;  E04CADC2BA74713B CRC64;

  Query Match             6.1%;  Score 148;  DB 5;  Length 1203;
  Best Local Similarity   20.3%;  Pred. No. 7.3;
  Matches  105;  Conservative   83;  Mismatches  258;  Indels   72;  Gaps   19;

Qy      12 VTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAV 71
Db     454 VTSTPTI-TSTQGAMASSSSNSPTSTQ-----AISSSTVTTVASSTIAPTSSQESSSIAS 507

Qy      72 DVHNAQITALIETRASRLHFEGETPATIADTFAKAEKL--DRLATT-----------TSG 118
Db     508 SSAPSQSTTVITVSTTATVSSGQSTSTFGTTIGQGSSFGSSTIQTTQGTSSFAPIPSTAG 567

Qy     119 ALRATPFAMAS-------LLQYMQPA--INKGDWLPAPLKPLTPLISGALSGAMDQVGTK 169
Db     568 SSSQTPGSMSSTGTTVGQMSSSFQPTAPTSLGTIMTTPGTSSIPSISTSVNSGSSTIGST 627

Qy     170 MMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTG-------VAVQTYSARNA 222
Db     628 VTQAPSSS---TSMGPSPSQSTAGSTMTSAPPVTSSSSANTGSTSSGTTVSVQTTQVSST 684

Qy     223 VRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDK 282
Db     685 TSPV---ASSSSQMTSTQQPSGSSSSIGSTVNQGSSS--VTTQPPASSRSTA-SQGSSSA 738

Qy     283 EPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLT 342
Db     739 QPIASSSTMGSTAGSSSPQPTASSTSVPSSTGATSSGSTVGSST--MGSTQSSLPSSTMT 796

Qy     343 QNGLALAGGFAGVGKLQEMAT--KNITDPATKAAVSQLTNLAGSAAVFA---------- 389
Db     797 NTG---STGSTVTNQLASSSTYGASSTEPIASSTANPGSSTSGQTAVTTQGSSSTQTNSN 853

Qy     390 GWTTAALTTDPAVKKA---ESFIQDTVKSTA---SSTTG--YVADQTVKLAKTVKDMG-- 439
Db     854 TGSTGSTVTNPPVSSSTSGSSSTQPIASSTANPGSSTSGTTVVTTQGSSSTQTNPNTGST 913

Qy     440 GEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPS 477
Db     914 GSTVTQPSAFSSTSASSSTQPIASSTTANPGSSTSGPT 951


RESULT 9
Q9F5K8
ID   Q9F5K8      PRELIMINARY;      PRT;  1363 AA.
AC   Q9F5K8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Flagellar hook-associated protein 1.
GN   FLGK.
OS   Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1063;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gonzalez-Pedrajo B., De La Mora J., Ballado T., Camarena L.,
RA   Dreyfus G.;
```

```
RT   "Isolation and Complementation of a Flagellar P-ring Mutant of
RT   Rhodobacter sphaeroides.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF317649; AAG31286.1; -.
DR   InterPro; IPR001444; Flag_bb_rod.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   Pfam; PF00460; flg_bb_rod; 1.
DR   PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ   SEQUENCE   1363 AA;  133154 MW;  AD7BB4856117808C CRC64;

  Query Match             6.1%;  Score 147.5;  DB 2;  Length 1363;
  Best Local Similarity   25.0%;  Pred. No. 9.2;
  Matches  132;  Conservative   47;  Mismatches  193;  Indels  157;  Gaps   26;

Qy      56 AHKLPPADSADGQAAVDVHN----------AQITALIETRASRLHFEGETPATIADTFAK 105
           |  |  |: | |   :   :           |:  ||: | |:     |  |  : |
Db     592 AGSLTAAEPAGGTIQIITRDGRHIAGAALTAEEAALLLTEAN-----GFLPGAVYD---- 642

Qy     106 AEKLDRLATTTSGA----LRATPFAMASL-------LQYMQPAINKGDWLP--APLKPLT 152
               |:| :||     | |  | | |       |    |       ||  : |  ||
Db     643 -------ASTLNGAGGTGFRGTGIAGAILPGERVLSLHPADPVAGSSGLLPPASALPSLT 695

Qy     153 -----------PLISGALSGAMDQ-------------VGTKMMDRATGDLHY----LSA 188
                     | :|| :  | |              |  :   | | | :     :
Db     696 LEAAGGLPLPVQLPAGASAAEMAQAINAFGAGIEAEARTGVTIEAPADGTLTFALTGTNL 755

Qy     184 SPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQ-GAVD 242
           || |:  |:|        :||        :||   ||   ||  |:| |   |||  |  |
Db     756 SPVRISGAVAGG---RMDALA-------LAVNAVSAATGVRAELSPDGARLLLVQDGGAD 805

Qy     243 LGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVL-GLKDKEPKAQLSEENDWLEAYKAI 301
           :|:    ||   ||    |    :      |  | | |  |    |: : |         :
Db     806 IGIV---GLRHTAGAAVTLQGTDAEGSPAGAPLTLSGTAD---SARFTGE---------L 850

Qy     302 KSASYSGAALNAGKRMAGLPLDMATDAM--GAVRSLVSASSLTQN-----------GLA 347
           : :| || : : |    |:  | | | |  | |   || :   |              ||:
Db     851 RLSSASGFSADLG----GVRQDAAVDPMSGGLVSRGVSGAGGVQTYGFTYDPAFDGAGLS 906

Qy     348 LAGGFAGVGKLQEMAT---KNITDPATKAAVSQLTNLAGS----------AAVFAGWTTA 394
             | ||  |  |   | : :|  |  | ||    :| :          ||   |   |
Db     907 ADGTFAQAGSAQYAMTVGNRTVTLDAAAAGVSDGAGVASALAALLRAEAPAATLTGGPVA 966

Qy     395 ALTTDPAVKKAESFIQDTVKSTASSTTGYVA------------DQT----VKLAKTVKD 437
           ||   ||  :: |   :    |  | | ||              | |    :: | ::
Db     967 AL---PADGRSVSVSYEGQSYTLRMTGGAVAVDGGEPGLLTAAFDATNRLVIQAAGSLDG 1023

Qy     438 MG-----GEAITHTGASLRNTVNNLRQRPAREA------DIEEGGTAAS 475
           |      | |     |:    |: |  :||  |      ||| |||  |
Db    1024 AGLRIESGAAAAFGLAAADAPVSTLTGQPADPAALPASFDIELGGTLYS 1072


RESULT 10
Q07290
ID   Q07290      PRELIMINARY;      PRT;  1822 AA.
AC   Q07290;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   EF protein.
GN   EPF*.
OS   Streptococcus suis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC   Streptococcaceae; Streptococcus.
OX   NCBI_TaxID=1307;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1890;
RX   MEDLINE=93328288; PubMed=8335363;
RA   Smith H.E., Reek F.H., Vecht U., Gielkens A.L.J., Smits M.A.;
RT   "Repeats in an extracellular protein of weakly pathogenic strains of
RT   Streptococcus suis type 2 are absent in pathogenic strains.";
RL   Infect. Immun. 61:3318-3326(1993).
DR   EMBL; X71880; CAA50714.1; -.
DR   TIGRFAMs; TIGR01168; YSIRK_signal; 1.
SQ   SEQUENCE   1822 AA;  192631 MW;  3838960C77641D7D CRC64;

  Query Match             6.1%;  Score 147.5;  DB 2;  Length 1822;
  Best Local Similarity   20.9%;  Pred. No. 14;
  Matches  106;  Conservative   78;  Mismatches  220;  Indels  103;  Gaps   19;

Qy      14 ATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDV 73
           | |:   | |  ||  ::|:   :  : :  :|:   |    |      |  || :
Db    1222 ARDAVELAKDKELAKEAIRTEEEEATKIVEKLAE---DTRKAIEDNPNLSDEDKQAEIKK 1278

Qy      74 HN---AQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTS----------GAL 120
             |:  | |   | :   | |    :||   ||::  ::|   :           | |
Db    1279 LTDAVAKTLATIRDNADKRTQEAEKAQALAD-LEKAKETQKIADKAAIDRLTILVKDGEL 1337

Qy     121 RATPFAMASLLQYMQPAINK---------------GDWLPAPLKPLTPLISGALSGAMDQ 165
            ||            | | ||              :   |  |  |  :   :: | |
Db    1338 EAT----------KQDAKNKIAKDAAAAKEAIASNPNLTDAEKKTFTDAVDAEVAKANDA 1387

Qy     166 VGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRT 225
           :          ||         |   :  ||  |::       |:|     :|     :|: |: :
Db    1388 ISA-----ATS-----PADVQKEEDAGVAAIAEDVLDAAKQDAKNKIAKDAAAAKEAIGS 1437

Qy     226 VLAPALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGAL---VLGLKDK 282
                 | :     |||  |: |    : |         ||  :     |:  ||    :
Db    1438 NPNLTDAEKKTFTDAVDAEVAKANDAISAA---TSPADVQKEEDAGVAAIAEDVLDAAKQ 1494

Qy     283 EPKAQLSEENDWLEAYKAIKSA-----SYSGAALNAGKRMAGLPLDMATDAMGAVRSLVS 337
           : | ::::|:|      | |||     : : |   : |:           ||||: |  | |
Db    1495 DAKNKIAKESD------AAKSAIDANPNLTDAEKESAKKAVDADAKAATDAIDASTSPVE 1548

Qy     338 ASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALT 397
           | |          ||| :    ::: | | : | :::     :|         |:
Db    1549 AQSAEDK---------GVGSI----AQDVLDAAKQDAKNKIAKEVAAA-------KEAID 1588

Qy     398 TDPAVKKAE-SFIQDTVKSTASSTTGYVADQTVKL-AKTVKDMGGEAITHTGASLRNTV- 454
             :| :  ||   :  | : | :||  :   |  : |:: :| |          |:|  |
Db    1589 ANPNLSDAEKEASKKAVDADAKATTDAIDASTSPVEAQSAEDKG-------VGSIRQDVL 1641

Qy     455 ----NNLRQRPAREADIEEGGTAASPS 477
                : : : |:|:|  :    |:|:
Db    1642 DAAKQDAKNKIAKESDAAKSAIDANPN 1668


RESULT 11
Q9Z5A4
ID   Q9Z5A4      PRELIMINARY;      PRT;  1156 AA.
AC   Q9Z5A4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative secreted protein.
GN   SCO6198 OR SC2G5.19.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Oliver K., Harris D.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
```

```
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2)..";
RL   Nature 417:141-147(2002).
CC   -!- COFACTOR: ZINC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC       FAMILY.
DR   EMBL; AL035478; CAB36606.1; -.
DR   InterPro; IPR002328; ADH_zinc.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   PROSITE; PS00059; ADH_ZINC; 1.
KW   Oxidoreductase; Zinc.
SQ   SEQUENCE   1156 AA;  116717 MW;  7046B814FB9517F8 CRC64;

  Query Match             6.1%;  Score 147;  DB 16;  Length 1156;
  Best Local Similarity   23.1%;  Pred. No. 7.8;
  Matches  122;  Conservative   64;  Mismatches  231;  Indels  112;  Gaps   21;

Qy      13 TATDSFRTA------SDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPAD--- 63
           || |: ||        |   :|::|  :: : ::    |:| :| | |   :|
Db      96 TAEDNRRTVLGAMALSGRGMATASAAALEAGEE----AVAAFL-DGGFKTAELQDLRFNT 150

Qy      64 -SADGQAAVDVHNAQITALIETRASRLH-FEGETPATIADTFAKAEKLDRLATTTSGALR 121
            |    : :| | |:::|          ||    |:    |: |   :  | |  :
Db     151 LSIQNSAGRGLKNAAQTALSDGTSDALSVFLLDTQFTARNTDERVEVFTILAN------- 203

Qy     122 ATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSG--AMDQVGTKMMDRATGDLH 179
             |   : :| | |:::|           ||    |:    |: |   :  | |  :
Db     204 ----ASPEVAKYAQRALDEG----------TP---SAIHWFLAIGQYIARARDEETATVD 246

Qy     180 YLSASPDR--------LHDAMAASVK-RHSPSLARQVLDTG----------VAVQTYSAR 220
             | |  :|            |:||| | : : : |::   |             ||    :||
Db     247 QLVAIVEREGKRAQITSDRAVAASDKAKEAAAKAKEAALTAAAEAEAAREDVAKSAAAAR 306

Qy     221 NAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGAL----- 275
             |       | |:| |||  :    | |   |  : |     :  :    |  ||
Db     307 KAANAAKGAASAARTAVQAS-----SAAHNAARRSAFAATAAAQAAATAGRAAALAYSAA 361

Qy     276 VLGLKD--KEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVR 333
           |   :|  |   |:|: |      | || |:|    ||  | :  |   :  |  |  :
Db     362 VAAARDASKTKAARLAAEGARNAAAKARKAAQALAAAQTATQAAAAAGISAAATARDSAA 421

Qy     334 S------LVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAA--VSQLTNLAGSA 385
           :        ||   |:  |:|  |     |          |||||     | | | ||
Db     422 AAQQAAVAAQASGAAQSEAAVARAAAAEADAQAAR-------ATKAANRAQSLANTAASA 474

Qy     386 AVFA-------------------GWTTAALTTDPAVKKAESFIQDTVKSTASSTTGYVAD 426
           |  |                    ||   |    ||:::  |:|   |:       |
Db     475 AAAARKAADSAAAHAEKAADAADAAADAAGEADDYANKAKAWAADSV--AAAELAAKAVD 532

Qy     427 QTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAAS 475
             :    ::   | : |       ::   |:  | ||:  |    |:
Db     533 DARAVEAAAREAEAEKLAH---DTEQSLAEAREMAAAEAEDREAARNAA 578


RESULT 12
Q93H87
ID   Q93H87      PRELIMINARY;      PRT;  6048 AA.
AC   Q93H87;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Modular polyketide synthase.
GN   PTEA1.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69303.1; -.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR004410; FabD.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00698; Acyl_transf; 4.
DR   Pfam; PF00109; ketoacyl-synt; 4.
DR   Pfam; PF02801; ketoacyl-synt_C; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   TIGRFAMs; TIGR00128; fabD; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_3.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_4.
KW   Phosphopantetheine.
SQ   SEQUENCE   6048 AA;  629420 MW;  F8673C2ED2694705 CRC64;

  Query Match             6.1%;  Score 147;  DB 2;  Length 6048;
  Best Local Similarity   22.1%;  Pred. No. 73;
  Matches  126;  Conservative   77;  Mismatches  204;  Indels  164;  Gaps   26;

Qy      13 TATDSFRTA-SD------ASLASSSVRSVSSDQ-------------------------QR 40
           ||| |    | :|       ||: | |:|: |::|
Db    5400 TATGSVTLAVADGAGDPVASVESLSLRAASAEQLPGAVGGDQLFGLEWTRFTLPSATDAL 5459

Qy      41 EINAIADYLTDHVFAAHKLP--------PADSADGQAAVD-VHN--AQITALI-----ET 84
           |  :||:  | :  :   |         |  :  |:    | ||:  ::: ||:     |
Db    5460 MIETVADF--DALRESEGTPDVVVVPVVPCPAGSGRETADRVHSTTSEVLALVQWWLAEE 5517

Qy      85 RASRLHFEGETPATIADT-------FAKAEKLDRL----ATTTSGALRATPFAMASLLQY 133
           || ||       |  :|         |::|  ||:     |  |  |:|  | |:||
Db    5518 RAGRLALV-TRPGDLAHAAVWGLVRSAQSENPDRIVLVEAEETDEAVRVLPAALAS--GE 5574

Qy     134 MQPAINKGDWLPAPLKPLTPLISGALSGAMD-QVGTKMMDRATGDLHYLSASPDRLHDAM 192
           | |:  |:      |   |   :|  :|  |  ||  ::  |:|              |:
Db    5575 PQFAVRDGEVFVPRLVKATRSATGTATGTPDFGVGPVLLTGASG--------------AL 5620

Qy     193 AASVKRHSPS---------LARQVLDTGVAVQ---TYSARNAVRTVLAPALASRPAVQGA 240
           |  ||  :           |:|:  :   ||:     :|  |       |  :| | || |
Db    5621 GGLVARHLVAEHGVRSLLLLSRRGAEAPGAVELEAELAAWGAEVRWAACDVADREAVSGM 5680

Qy     241 VD-LGVSMAGGLAANAG-FGNRLLSVQSRDHQR--------------------GGAL--- 275
           :: ||      :   ||  |: :::  : :  |                     | |
Db    5681 LNGLGEQSLSAVVHTAGVLGDGIVASLTPERMREVFRPKVDAVLNLHECTRDMGLAAFVV 5740

Qy     276 ---VLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMAGLP----------- 321
           | |:     :|  :  | :|:|:           :| :|  |||
Db    5741 FSSVAGMVGSAGQASYAAANSFLDAF-------------SAHRRREGLPAISLAWGVWEQ 5787

Qy     322 ----------LDMATDAMGAVRSLVSASSLTQNGLALAGGFA-------GVGKLQEMATK 364
                     | |  |   |  |     |    |||   |          | |:
Db    5788 SGAMTDGLVEADRARMARSGVLPLPPEEGLRLFDAALASDEAVLAPVRIDTGALRAGEAP 5847

Qy     365 NITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASSTTGYV 424
```

```
Db    5848 PVLRALVPAAAAARRTARA-AAAPSASSLAERLEEMPEAER-EKTVLDLVRAEVAAGLGHA 5905

Qy     425 ADQTVKLAKTVKDMGGEAITHTGASLRNTVN 455
Db    5906 SDRTVRPEHAFQDLGFDSL--TAVELRNRLN 5934
```

RESULT 13
Q8XDQ4
ID Q8XDQ4 PRELIMINARY; PRT; 973 AA.
AC Q8XDQ4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative membrane protein of prophage CP-933X (Putative tail fiber
DE protein).
GN Z1918 OR ECS1650.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005333; AAG56007.1; -.
DR EMBL; AP002555; BAB35073.1; ALT_INIT.
DR InterPro; IPR000104; Antifreeze_1.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR005003; Phage_fiber.
DR InterPro; IPR005068; Phage_fiber_2.
DR Pfam; PF03335; Phage_fiber; 6.
DR Pfam; PF03406; Phage_fiber_2; 3.
DR PRINTS; PR00308; ANTIFREEZEI.
KW Complete proteome.
SQ SEQUENCE 973 AA; 96316 MW; 776580D2A87E1B36 CRC64;

Query Match 6.0%; Score 145.5; DB 16; Length 973;
Best Local Similarity 22.5%; Pred. No. 7.4;
Matches 107; Conservative 52; Mismatches 232; Indels 85; Gaps 15;

```
Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQI 78
Db     117 RNASAVAQNTAAAKKSASDASTSAREAATHATD---AADSARAASTSAGQAASSAQSASS 173

Qy      79 TALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQPAI 138
Db     174 SA----------GTASTKATEASKSAAAAESSKSAAATSAGAAKTSETNAAVSQ--QSAA 221

Qy     139 NKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAA--SV 196
Db     222 TSA-------STATTKASEAASSARDASASKEAAKSS-ETSAASSASSAAASSATAAGNSA 273
```

```
Qy     197 K---------RHSPSLARQVLDTGVAVQTYSA--RNAVRTVLAPALASRPAVQGAVDLGV 245
Db     274 KAAKTSETNAKSSETAAEQSASAAAGSKTAAALSASAASTSAGQASASATAAGKSAESAA 333

Qy     246 SMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSAS 305
Db     334 SSASTATTKAGEATEQASAAASSASAAKTSETNAKASETSAESSKTAAASSASSAASSAS 393

Qy     306 YSGAA-------LNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKL 358
Db     394 SASASKDEATRQASAAKSSATTASTKATEAAGSATAAAQSKSTAESAATRA--------- 444

Qy     359 QEMATKNITDPA----------TKAAVSQLTNLAGSAAVFAGWTTAALTTDP-AVKKAES 407
Db     445 -ETAAKRAEDIASAVALEDASTTKKGIVQLSSATNS-------TSESLAATPKAVKAAYE 496

Qy     408 F------IQDTVK--------STASSTTGYVADQTVKLAKTVKDMGGEAITHTGAS 449
Db     497 LANGKYTAQDATTAQKGIVQLSNATNSTSEMLAATPKSVKAAYDLANGKYTAQDAT 552
```

RESULT 14
Q98M03
ID Q98M03 PRELIMINARY; PRT; 2124 AA.
AC Q98M03;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Kinesin-like protein.
GN MLR0796.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002995; BAB48310.1; -.
KW Complete proteome.
SQ SEQUENCE 2124 AA; 229253 MW; 3457A2B462EFED01 CRC64;

Query Match 6.0%; Score 144.5; DB 16; Length 2124;
Best Local Similarity 21.2%; Pred. No. 24;
Matches 108; Conservative 86; Mismatches 191; Indels 125; Gaps 22;

```
Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQA--------- 69
Db     405 RTTGMVSLLGGAARTLNSEFEASLNGIERTLAER----------GQALISEFQTRA 450

Qy      70 -AVDVHNAQITALIETRASRLHFE-GETPATIADTFAK-----AEKLDR----------- 111
Db     451 EALDTGTQKLNAALEARARQINETLVERARE IAHTFAESKDTLAAMIDQGKTQIGADMAD 510

Qy     112 LATTTSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMM 171
Db     511 IVTSTSSMLEARASDFAGRMEAARHVVSRS--FDSDIQRLADARVG-IEEAVENHSRK-- 565

Qy     172 DRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPAL 231
Db     566 ---------LSESRDRMAAAMQADLEKFAES--RDGIDAAVTNQVQKLAEG-RSLIARAL 613

Qy     232 ASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEE 291
Db     614 EE--------DLRKVNESRAAIDASLGSHLERLEE------GRNRLSL-------ALNED 652
```

```
Qy     292 NDWLEAYKAIKSASYSGAALNAGKRMAG-------LPLDMA--TDAMGAVRSLVS----- 337
           :  |   : |      :|   : ||   |       |  |:  :|:  ::  ||:
Db     653 SGKLVQARTIIDEMVAG---HVGKLAEGRNILSRALEADLGKLSDSRASIDGLVAGQVEK 709

Qy     338 -ASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAAL 396
            |          ||    ||:  | |: :  : :  ::   |   :|:|           |
Db     710 IAEGRAVLAKALENDIAGIKSLIEVHSAKLVEDRSQLGRSLEDDLSG---------IRGL 760

Qy     397 TTDPAVKKA--ESFIQDTVKS-------TASSTTGYVADQTVKLAKTVKDMGGEAITHTG 447
           | :|| |    | :  |:::       : ||  | || |  |||:  :|:   |:
Db     761 LDDHSVKLANDRSLLSQTLEADLAKLAESRSSIDGLVAGQVEKLAEG-RDILKRALESDL 819

Qy     448 ASLRNTVNNLRQRPAR---------EADIE 468
           ::::  :::  :| |           |||::
Db     820 STIKGVISSQSERLAEDRGQLSRVLEADLQ 849
```

```
RESULT 15
P91365
ID   P91365      PRELIMINARY;      PRT;  2232 AA.
AC   P91365;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   K06A9.1 protein.
GN   K06A9.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Geisel C., Gattung S.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS: TWO FORMS (A AND B) MAY BE PRODUCED BY
CC       ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS THAT
CC       OF THE A FORM.
DR   EMBL; U80846; AAC70889.1; -.
DR   EMBL; U80846; AAC70890.1; -.
KW   Alternative splicing.
FT   VARSPLIC    842    866       VASSSPAPSTSQNPNPSTSSGSSMI -> LATTSAPKPSVT
FT                                CLFMYD (IN ISOFORM B).
FT   VARSPLIC    870   1051       PYPSQSTSPVESSTTPSPGSPGTTLTSTSPSPSQSTTIGST
FT                                QGSTSPGISTTSEEMTSQGSTQTPGSTGSTVTQPSTVSDST
FT                                SSGSTVTVGSTEGSSSPIPSTSQNTNPSTSSGSSMSTQTPQ
FT                                SSQSTSPVESSTSGATSSSGSPGTTLTSISPSPSPSSTIGS
FT                                SQGSTSPVVSTISQGSTE -> KEIDQTAINTYKTYFNFAL
FT                                LVASKLNNESILTGYIDNFGYSAGLNDHQYYPTDDYNGIKS
FT                                VPFPIDGTDDDIDLDLKDVDKSLATADWTPPVADQTCMIFI
FT                                SAAPEDEYGGTTIKSTYTYFETVVGVLVGGAKSIPGLSIDK
FT                                NIVITNNTMNDRDASAVVSKLLELLPTA (IN ISOFORM
FT                                B).
SQ   SEQUENCE   2232 AA;  213840 MW;  08D69FA638E14CC8 CRC64;

  Query Match             6.0%;  Score 144.5;  DB 5;  Length 2232;
  Best Local Similarity   18.7%;  Pred. No. 26;
  Matches   99;  Conservative   81;  Mismatches  261;  Indels   89;  Gaps   11;

Qy       3 INRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPP- 61
           |   | ||  | : |  : |  :: ||   :  |  :   ::|:   |     :  :|
Db    1306 IGSTVTQPS-TVSGSNSSGSTVTIGSSEASTSGSSFKTSPSSISPVPT-----SSPIPST 1359

Qy      62 --ADSADGQAAVDVHNAQITALI----------------------ETRASRLHFEGETP 96
             | |  |     || :   |:|                      ::||      :|
Db    1360 TFASSTSGSTISDVSSVSTTSLAPLSSSLPSTVPSSTQSFSSTSEGSSKASSSPVPSQTS 1419

Qy      97 ATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQPAINK--GDWLPAPLKPLTPL 154
           :|  :     |    |::| ||: : |  : ||      |: |   |  :       : :
Db    1420 STPTNPTGSTESSTLLSSTISGSTQHTTMSKAS-SGSTSPSTNSQTGSTVTMGSSSTSGV 1478

Qy     155 ISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAA------------SVKRHSP 201
```

```
           : : |    |: |     |   :   :|||     | ::              |  . |
Db    1479 STSSASSTQPQMSTSQGSSAGSTVASSTASPAASSTAPSSTGTMSSTSSGTVGSTISESS 1538

Qy     202 SLARQVLDTGVAV-----------------------QTYSARNAVRTVLAPALASRPA 236
           : |      ||  |                       |  || : | :  |  :::
Db    1539 TTASASSQTGSTVTMGSSSTSGVSTSSASSTQPQMSTSQGSSAGSTVASSTAGLVSTSTV 1598

Qy     237 VQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLE 296
                :| : :| : :     :   |  |   | |  : :|       :  |  :   :
Db    1599 PSSTGTMGSTSSGTVGSTISESSTTASASS---QTGSTVTMGSSSTSGVSTSSASSTQPQ 1655

Qy     297 AYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVG 356
             :  |:: |  | :    ::   :  :|  ||:  |    |:::::  | |   :  |
Db    1656 MSTSQGSSAGSTVASSTTGLVSTSTVPSSTGTMGSTSSGTVGSTISESSTA-ASASSQTG 1714

Qy     357 KLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKST 416
             |  : :  :| :| |    :: |     || |  : |  ||          |: ||
Db    1715 STVTMGSSSTSGVSTSSASSGQPQMSTSQGSSAGSTVVSSTASPAASSTAPSSTGTMSST 1774

Qy     417 ASSTTGYVADQTVKLAKTVKDMG-----------GEAITHTGASLRNTV 454
           :| | |    |:   | |    |           :  |  |:|: :||
Db    1775 SSGTVGSTMSQSSTAASTTSHTGSTVTLGSSSTSSNQMSTSQGSSVGSTV 1824
```

Search completed: January 2, 2003, 15:20:32
Job time : 42.9692 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:13:55 ; Search time 34.4645 Seconds
(without alignments)
1879.027 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHINRRVQQPPVTATDSFRT.....IEEGGTAASPSEIPFRPMRS 486

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2412	100.0	486	22	AAE12573	Pseudomonas syring
2	2412	100.0	486	22	AAB67677	Amino acid sequenc
3	1869.5	77.5	487	22	AAE12603	Pseudomonas syring
4	146	6.1	1463	23	AAE20110	Lactobacillus rham
5	144.5	6.0	1822	13	AAR27745	Extracellular fact
6	143	5.9	600	20	AAY32068	Mycobacterium tube
7	143	5.9	600	23	AAU74597	Antigenic fusion p
8	142.5	5.9	1134	22	AAM78810	Human protein SEQ
9	142.5	5.9	1144	22	AAM39293	Human polypeptide
10	142.5	5.9	1165	22	AAM39312	Human polypeptide
11	139.5	5.8	1177	22	AAM79794	Human protein SEQ
12	139.5	5.8	1177	22	AAM41079	Human polypeptide
13	139.5	5.8	1177	22	AAM41098	Human polypeptide
14	135	5.6	723	21	AAY75477	Neisseria meningit
15	133.5	5.5	729	23	AAE17572	Mycobacterium spec
16	132	5.5	1721	19	AAW52847	A. mediterranei ri
17	132	5.5	1787	23	ABB49791	Listeria monocytog
18	131.5	5.5	2086	22	AAU34143	Staphylococcus aur
19	131.5	5.5	3596	21	AAY87407	Bordetella pertuss
20	131.5	5.5	5795	22	AAU37017	Staphylococcus aur
21	130.5	5.4	1996	18	AAW22607	Platenolide syntha
22	130.5	5.4	1996	18	AAW23717	Platenolide syntha
23	130.5	5.4	3647	11	AAR05041	Filamentous haemag
24	129.5	5.4	1201	22	ABB61629	Drosophila melanog
25	129.5	5.4	2344	22	AAU37120	Staphylococcus aur
26	129	5.3	2434	22	AAU34339	Staphylococcus aur
27	129	5.3	6281	22	AAU37403	Staphylococcus aur
28	128.5	5.3	1026	15	AAR48993	rsaA S-lyaer prote
29	128.5	5.3	1026	17	AAR94014	Caulobacter S-laye
30	128.5	5.3	3257	22	ABB67502	Drosophila melanog
31	127.5	5.3	596	20	AAY32070	Mycobacterium tube
32	127.5	5.3	596	23	AAE17574	Mycobacterium spec
33	127.5	5.3	599	23	AAU74599	Antigenic fusion p
34	127.5	5.3	729	22	AAO22142	Ra12-H9-32A fusion
35	127.5	5.3	729	23	AAE17573	Mycobacterium spec
36	127.5	5.3	2541	21	AAB41087	Human ORFX ORF851
37	127	5.3	676	21	AAY59271	Human huntingtin-i
38	127	5.3	2123	22	AAE00701	Moraxella catarrha
39	126.5	5.2	669	20	AAY00141	Enterococcus faeca
40	126.5	5.2	669	23	ABP43360	E faecalis EF069 a
41	126.5	5.2	1026	18	AAW37490	Caulobacter cresce
42	126.5	5.2	1026	21	AAY44757	Caulobacter cresce
43	126.5	5.2	1638	20	AAY00138	Enterococcus faeca
44	126.5	5.2	1638	20	AAY00140	Enterococcus faeca
45	126.5	5.2	1638	20	AAY00142	Enterococcus faeca

ALIGNMENTS

RESULT 1
AAE12573
ID AAE12573 standard; Protein; 486 AA.
XX
AC AAE12573;
XX
DT 03-JAN-2002 (first entry)
XX
DE Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 encoded protein.
XX
KW Conserved Effector Loci; CEL; cytostatic; antibacterial; gene therapy;
KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
KW eukaryotic cell death; cancer.
XX
OS Pseudomonas syringae.
XX
PN WO200175066-A2.
XX
PD 11-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA (UYNE-) UNIV NEBRASKA.
XX
PI Collmer A, Alfano JR, Charkowski AO;
XX
DR WPI; 2001-639361/73.

DR N-PSDB; AAD20408.
XX
PT New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
PT genomic sequences, for imparting disease resistance to plants -
XX
PS Claim 8; Page 21-23; 217pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant
CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC Pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targetted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targetted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is
CC Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 protein.
XX
SQ Sequence 486 AA;

Query Match 100.0%; Score 2412; DB 22; Length 486;
Best Local Similarity 100.0%; Pred. No. 7e-193;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60
Db       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60

Qy      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120
Db      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120

Qy     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180
Db     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180

Qy     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240
Db     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240

Qy     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300
Db     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300

Qy     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360
Db     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360

Qy     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420
Db     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420

Qy     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEIP 480
Db     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEIP 480

Qy     481 FRPMRS 486
           ||||||
```

Db 481 FRPMRS 486

RESULT 2
AAB67677
ID AAB67677 standard; Protein; 486 AA.
XX
AC AAB67677;
XX
DT 11-JUN-2001 (first entry)
XX
DE Amino acid sequence of a HopPtoA protein of Pseudomonas syringae.
XX
KW HIV; tat protein; effector protein; transduction domain; HopPtoA protein.
XX
OS Pseudomonas syringae.
XX
PN WO200119393-A1.
XX
PD 22-MAR-2001.
XX
PF 13-SEP-2000; 2000WO-US24977.
XX
PR 13-SEP-1999; 99US-0153507.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Collmer A, Beer SV;
XX
DR WPI; 2001-257850/26.
DR N-PSDB; AAF55683.
XX
PT Delivering effector proteins into target cell for use in protein
PT therapy, involves introducing effector protein fused to protein
PT transduction domain of human immunodeficiency virus TAT protein, into
PT target cells -
XX
PS Example 2; Page 23; 43pp; English.
XX
CC The present sequence represents a HopPtoA protein of Pseudomonas syringae
CC pv. tomato DC300 CEL.. This is an effector protein, which is used in
CC the method of the invention. The specification describes a method for
CC delivering effector proteins into a target cell. The method comprises
CC introducing an effector protein fused to a protein transduction domain
CC of a human immunodeficiency virus (HIV) tat protein into the target
CC cell. The method is used for delivering a heterologous effector protein
CC such as an effector protein produced by bacterial plant pathogen, animal
CC pathogen or a rhizosphere bacteria, or a protein secreted and/or
CC delivered into eukaryotic cells by a Type III secretion system or a
CC hypersensitive response elicitor, an avirulence protein, a
CC hypersensitive response and pathogenicity-dependent outer protein, a
CC virulence protein or a pathogenicity protein, into an eukaryotic cell.
CC The method is useful for delivering effector proteins for use in
CC pharmaceutical, insecticidal, fungicide, herbicide and other
CC applications.
XX
SQ Sequence 486 AA;

Query Match 100.0%; Score 2412; DB 22; Length 486;
Best Local Similarity 100.0%; Pred. No. 7e-193;
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60
Db       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60

Qy      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120
Db      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120

Qy     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180
Db     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180
```

Not double pat, 6 lc dft inventors. Isolate

```
Qy     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240

Qy     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300

Qy     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360

Qy     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420

Qy     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEIP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEIP 480

Qy     481 FRPMRS 486
           ||||||
Db     481 FRPMRS 486
```

RESULT 3
AAE12603
ID AAE12603 standard; Protein; 487 AA.
XX
AC AAE12603;
XX
DT 03-JAN-2002 (first entry)
XX
DE Pseudomonas syringae pv. tomato strain DC3000 HopPtoA2 protein.
XX
KW Conserved Effector Loci; CEL; cytostatic; antibacterial; gene therapy;
KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
KW eukaryotic cell death; cancer.
XX
OS Pseudomonas syringae.
XX
PN WO200175066-A2.
XX
PD 11-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA (UYNE-) UNIV NEBRASKA.
XX
PI Collmer A, Alfano JR, Charkowski AO;
XX
DR WPI; 2001-639361/73.
DR N-PSDB; AAD20438.
XX
PT New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
PT genomic sequences, for imparting disease resistance to plants -
XX
PS Claim 8; Page 65-66; 217pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant
CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC Pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targetted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targetted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is
CC Pseudomonas syringae pv. syringae HopPsyA homolog protein.
XX
SQ Sequence 487 AA;

```
  Query Match             77.5%;  Score 1869.5;  DB 22;  Length 487;
  Best Local Similarity   78.4%;  Pred. No. 1.4e-147;
  Matches  381;  Conservative   33;  Mismatches   71;  Indels    1;  Gaps    1;

Qy       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60
           ||||:  ||||  | :||||||||||||||||||:  |:: || |||  ||||||:
Db       1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60

Qy      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120
             | | : |   || || ||:||||:||: ||||||||| |||||| | |
Db      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120

Qy     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180
             |||| ||:||||||||||||||  |||||||||||||||||||||||||| |||||
Db     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180

Qy     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240
           || |||:|||||| |||||||:| |||:| |:||||:|| | ||||||||||||:||||
Db     181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240

Qy     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300
           || ||| |||| ||||||:|:|||||||  |||| |||:||||||| |||| |||:||||
Db     241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300

Qy     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360
           |||||||||||||||||||||:||| : |||||||||:|||:||||||||:||| |||:
Db     301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360

Qy     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420
           |||||||| |||||||||:|| ||  |||||||  | |||||||||||||| ||||||||
Db     361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420

Qy     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAA-SPSEI 479
           | |||||||||||||||| |||||: |||||:||||||  | | ||||||||  :| | ||
Db     421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480

Qy     480 PFRPMR 485
           ||:  |
Db     481 PFQLRR 486
```

RESULT 4
AAE20110
ID AAE20110 standard; Protein; 1463 AA.
XX
AC AAE20110;
XX
DT 18-JUN-2002 (first entry)
XX
DE Lactobacillus rhamnosus outer membrane protein rompA.
XX
KW Enzyme; flavour; aroma; texture; nutritional; dairy manufacture; therapy;

```
KW   fermentation process; anti-infection; rotavirus infection; heart disease;
KW   infantile diarrhoea; lactose digestion; anti-cancer; autoimmune disorder;
KW   anti-mutagenesis; immune system modulation; allergy; Helicobacter pylori;
KW   antihypertensive effect; urogenical infection; hepatic encephalopathy;
KW   bowel syndrome; endocarditis; transgenic microbe; outer membrane protein;
KW   rompA.
XX
OS   Lactobacillus rhamnosus HN001.
XX
PN   WO200212506-A1.
XX
PD   14-FEB-2002.
XX
PF   08-AUG-2001; 2001WO-NZ00160.
XX
PR   08-AUG-2000; 2000US-0634238.
PR   28-NOV-2000; 2000US-0724623.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (VIAL-) VIALACTIA BIOSCIENCE NZ LTD.
XX
PI   Glenn M, Havukkala IJ, Bloksberg LN, Lubbers MW, Dekker J;
PI   Christensson AC, Holland R, O'toole PW, Reid JR, Coolbear T;
XX
DR   WPI; 2002-241760/29.
DR   N-PSDB; AAD31881.
XX
PT   New polynucleotides and polypeptides from Lactobacillus rhamnosus,
PT   useful in e.g. improving the flavor, aroma, texture and health-related
PT   benefits of milk-derived products, or in increasing properties of
PT   microbes -
XX
PS   Claim 2; Fig 64; 257pp; English.
XX
CC   The present invention relates to a new isolated polynucleotide comprising
CC   a sequence present in Lactobacillus rhamnosus strain HN001 and encoding a
CC   polypeptide capable of modifying the flavour, aroma, texture, nutritional
CC   and health benefits of milk-derived products, and/or survivability of
CC   microbes in dairy manufacturing processes. The polynucleotides are useful
CC   for improving the properties of microbes used in the manufacture of milk-
CC   derived products such as cheeses, yogurt, fermented milk products, sour
CC   milks and buttermilk; in modifying the flavour, aroma, texture and health
CC   -related benefits of milk-derived products and in increasing the survival
CC   of microbes during industrial fermentation processes. The bacteria may be
CC   used to increase resistance to enteric pathogens and anti-infection
CC   activity, including treatment of rotavirus infection and infantile
CC   diarrhoea; aid in lactose digestion; as anti-cancer and anti-mutagenesis;
CC   liver cancer reduction; reduction of small bowel bacterial overgrowth;
CC   immune system modulation and treatment of autoimmune disorders and
CC   allergies; treatment of allergic responses to foods; reduction of blood
CC   lipids and prevention of heart disease; antihypertensive effect;
CC   prevention and treatment of urogenical infections, Helicobacter pylori,
CC   or hepatic encephalopathy; treatment of inflammatory bowel disorder and
CC   irritable bowel syndrome; modulation of endocarditis; and for improved
CC   protein and carbohydrate utilization and conversion. The transgenic
CC   microbial population can be administered to a mammal as an anti-
CC   carcinogenic agent. The present sequence is Lactobacillus rhamnosus
CC   outer membrane protein rompA.
XX
SQ   Sequence   1463 AA;

  Query Match             6.1%;  Score 146;  DB 23;  Length 1463;
  Best Local Similarity   22.5%;  Pred. No. 0.0056;
  Matches  112;  Conservative   72;  Mismatches  219;  Indels   94;  Gaps   20;

Qy      14 ATDSFRTASDASLASSSVR---SVSSD--QQREINAIADYL------TDHVFAAHKLPPA 62
           |::: : :|:|| |:|:     | :||  :| :  | || :      | :  |: 
Db     996 ASEAAKASSNASAATSAAAVGFSAASDASEQAKTAASADVVASSAASTANSNASAAASAT 1055

Qy      63 DSADGQAAVDVHNAQITALIETR-ASRLHFEGETPATIADTFAK-----AEKLDRLATTT 116
             : | :||    :|   |    : |  :  |  : |   |: |      |   |: |:
Db    1056 KAGDSKAAAGFSSAASAAASSAKGAEAVASEAASAAASDDSVASSAASAAAGFDKAASAA 1115

Qy     117 SGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATG 176
             ||  :   | ||       |  :|              |  | :  : |      :|
Db    1116 EGAASSAASAAAS------SAAAQG-------------TRGGASSSASEAG----QASTA 1152

Qy     177 DLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPA 236
             | ||:     :  | |    : | | :   || |    |  :|     | : |||
Db    1153 TSVYASAASSSASE--AGSYAHQAGSSASEA--TGHASSATSQASA-----ASSAASRYP 1203

Qy     237 VQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLE 296
              :  ||:|   |: |       |  |                  :| ::   | :
Db    1204 SDSGIQSDVSIASSAASTAS------SAASAAQSEASTASSAASHASEQASIASSEDVVS 1257

Qy     297 AYKAIKSASYSGAALNAGK----RMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGF 352
             :  |  ::| : || :| |      ||:    |: |  : :|  | :|   :  |
Db    1258 SSAASVASSAASAASSAAKAGNSSAAGIYSHAASAAASSAKSAESQASSAASAAA----- 1312

Qy     353 AGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAV-------KKA 405
                 :  ::   |  ||:|     : :| | :  |||| :::  ::          |
Db    1313 ---------SDDSVASSAASAALSDDAKASSAADVASSATTAAISSATSLADQSATGSTA 1363

Qy     406 ESFIQDTVKSTASSTTGYV-ADQT--------VKLAKTVKDMGGEAITHTGASLRNTVNN 456
             | |   : ||   ||| : : ||          |  :|     |:  | || ||  | :
Db    1364 GSHI---LPSTGGETTGSIPSGQTPTQTKPTQTKPTQTKPTQAGQT-TQTG-SLPQTDHA 1418

Qy     457 LRQRPAREADIEEGGTA 473
             |    :  |  | ||:
Db    1419 GRHMLPQTGDDAESGTS 1435


RESULT 5
AAR27745
ID   AAR27745 standard; Protein; 1822 AA.
XX
AC   AAR27745;
XX
DT   04-MAR-1993  (first entry)
XX
DE   Extracellular factor related protein.
XX
KW   EF*; detection; prevention; screening; diagnostic.
XX
OS   Streptococcus suis type II (non-pathogenic).
XX
FH   Key             Location/Qualifiers
FT   Peptide          1..46
FT                    /note= "signal peptide"
FT   Peptide          47..1822
FT                    /note= "mature peptide"
FT   Region           858..861
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
FT   Region           934..937
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
FT   Region           999..1002
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
FT   Region           1075..1078
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
FT   Region           1264..1267
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
FT   Region           1362..1365
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
FT   Region           1438..1441
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
FT   Region           1514..1517
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
FT   Region           1590..1593
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
FT   Region           1666..1669
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
FT   Region           1741..1744
FT                    /note= "repetitive Asn-Pro-Asn-Leu sequence"
XX
PN   WO9216630-A.
XX
```

PD 01-OCT-1992.
XX
PF 19-MAR-1992; 92WO-NL00054.
XX
PR 21-MAR-1991; 91NL-0000510.
XX
PA (DIER-) CENT DIERGENEESKUNDIG INST.
XX
PI Smith HE, Vecht U;
XX
DR WPI; 1992-349215/42.
DR N-PSDB; AAQ29471.
XX
PT Deoxyribonucleic acid encoding virulence characteristic of
PT Streptococcus suis - useful for antibody and polypeptide for
PT diagnosing and preventing infections in pigs and humans
XX
PS Claim 9; Fig 1b; 86pp; English.
XX
CC The sequence is that of the extracellular factor related protein
CC from Streptococcus suis type II (non-pathogenic) which allows the
CC detection and the prevention of infections by S. suis in a more
CC effective manner than was previously possible. It facilitates
CC screening of e.g. pigs and elimination of infected and carrier pigs
CC can then be carried out. The new diagnostic tests can distinguish
CC between avirulent and virulent strains. It may be used in the prodn.
CC of a vaccine. See also AAR27744 and AAR27746.
XX
SQ Sequence 1822 AA;

Query Match 6.0%; Score 144.5; DB 13; Length 1822;
Best Local Similarity 21.2%; Pred. No. 0.01;
Matches 109; Conservative 81; Mismatches 208; Indels 115; Gaps 22;

```
Qy      14 ATDSFRTASDASLASSSVRSVSSDQQREINAIAD----------------------YLTD 51
           | |:   | |   |    ::|:   :  : :  :|:                       |||
Db    1222 ARDAVELAKDKELGKEAIRTEEEEATKIVEKLAEDTRKAIEDNPNLSDEDKQAEIKKLTD 1281

Qy      52 HVFAAHKLPPADSAD--------GQAAVDVHNAQITALIETRAS--RLHF---EGETPAT 98
             | |        |:||         ||  |:  |: |  |  :|:  ||     :||  ||
Db    1282 AV-AKTLATMRDNADKRTQEAEKAQALADLEKAKETQKIADKAAIDRLTILVKDGELEAT 1340

Qy      99 IADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGA 158
             |  ||  |: : |      |: : |                |  |   |  |  |  :
Db    1341 KQD--AKT-KIAKDAAAAKEAIASNP-------------NLTD---AEKKTFTDAVDAE 1380

Qy     159 LSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYS 218
           :: | | :       ||       |   :  ||  |::       |:|     :|    :
Db    1381 VAKANDAISA-----ATS-----PADVQKEEDAGVAAIAEDVLDAAKQDAKNKIAKDAAA 1430

Qy     219 ARNAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGAL--- 275
           |: |: :       | :     |||  |: |    : |          ||  :     |:
Db    1431 AKEAIGSNPNLTDAEKKTFTDAVDAEVAKANDAISAA---TSPADVQKEEDAGVAAIAED 1487

Qy     276 VLGLKDKEPKAQLSEENDWLEAYKAIKSA-----SYSGAALNAGKRMAGLPLDMATDAMG 330
           ||    :: | ::::|:|       | |||     : : |   : |:        ||||:
Db    1488 VLDAAKQDAKNKIAKESD------AAKSAIDANPNLTDAEKESAKKAVDADAKAATDAID 1541

Qy     331 AVRSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAG 390
           |  | | | |              ||| :      ::: | | : | :::     :|
Db    1542 ASTSPVEAQSAEDK---------GVGSI----AQDVLDAAKQDAKNKIAKEVAAA----- 1583

Qy     391 WTTAALTTDPAVKKAE-SFIQDTVKSTASSTTGYVADQTVKL-AKTVKDMGGEAITHTGA 448
             |:  :| :  ||     :  | : | :||  :    |  : |:: :| |       |:
Db    1584 --KEAIDANPNLSDAEKEASKKAVDADAKATTDAIDASTSPVEAQSAEDKG------VGS 1635

Qy     449 SLRNTVNNLRQ----RPAREADIEEGGTAASPS 477
             :: ::  :|    : |:|:|  :    |:|:
Db    1636 IAQDVLDAAKQDAKNKIAKESDAAKSAIDANPN 1668
```

RESULT 6
AAY32068

ID AAY32068 standard; Protein; 600 AA.
XX
AC AAY32068;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mycobacterium tuberculosis antigen fusion protein Mtb61f.
XX
KW Tuberculosis; antigen; fusion protein; Mtb61f; TbH9; DPV; MTI;
KW diagnosis; therapy; vaccine; immunogen.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9951748-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07717.
XX
PR 07-APR-1998; 98US-0056556.
PR 30-DEC-1998; 98US-0223040.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Alderson M, Campos-Neto A;
XX
DR WPI; 1999-601610/51.
DR N-PSDB; AAZ20203.
XX
PT New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis -
XX
PS Claim 1; Fig 10A-B; 83pp; English.
XX
CC This sequence represents a recombinant Mycobacterium tuberculosis
CC tri-antigen fusion protein, termed Mtb61f, composed of the antigens
CC TbH9, DPV and MTI. The fusion protein is expressed in host cells
CC using a vector carrying a polynucleotide (see AAZ20203) comprising
CC the coding sequences for the 3 antigens. The invention provides
CC fusion proteins (see AAY32059-71) containing at least 2 M.
CC tuberculosis antigens. The new fusion proteins and polynucleotides
CC encoding them are useful as vaccines for preventing tuberculosis
CC (claimed), for diagnosis (via in vitro assays or intradermal skin
CC tests for detection of anti-M. tuberculosis antibodies), monitoring
CC of disease progression, and treatment of tuberculosis. They are
CC more effective immunogens than mixtures of the individual protein
CC components.
XX
SQ Sequence 600 AA;

Query Match 5.9%; Score 143; DB 20; Length 600;
Best Local Similarity 23.0%; Pred. No. 0.0029;
Matches 117; Conservative 58; Mismatches 165; Indels 168; Gaps 25;

```
Qy      20 TASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQA---------- 69
           ||  | | :: ||           | | | | :      :||    |: :|
Db      88 TAGQAELTAAQVRV---------AAAAYETAYGLT---VPPPVIAENRAELMILIATNL 134

Qy      70 ------AVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL--- 120
                 |: |: |:   :     |: :       ||    |     |:   :  |::| |
Db    135 LGQNTPAIAVNEAEYGEMWAQDAAAMFGYAAATATATATLLPFEEAPEM--TSAGGLLEQ 192

Qy     121 -----RATPFAMAS-LLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVG------T 168
                 |:  | |: |:  :  |: :   |  | : ||  |  | |      |       :
Db    193 AAAVEEASDTAAANQLMNNVPQALQQ---LAQPTQGTTP--SSKLGGLWKTVSPHRSPIS 247

Qy     169 KMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLA 228
             |:  |   :  :: |   : : ::: :|  :|: | |       |||| :|:|  ||
Db    248 NMVSMANNHMS-MTNSGVSMTNTLSSMLKGFAPAAAAQ------AVQT-AAQNGVR---- 295

Qy     229 PALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQL 288
             |::|  :  |:  ||    ||:||| |   | || |                    |
Db    296 -AMSSLGSSLGSSGLG----GGVAANLG---RAASVGS--------------------L 326
```

```
Qy    289 SEENDWLEAYKAIKSAS--------YSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASS 340
Db    327 SVPQAWAAANQAVTPAARALPLTSLTSAAERGPGQMLGGLPVG----------------- 369

Qy    341 LTQNGLALAGGFAGV----------------GKL--------------QEMATKNITDP- 369
Db    370 --QMGARAGGGLSGVLRVPPRPYVMPHSPAAGKLDPVDAVINTTCNYGQVVAALNATDPG 427

Qy    370 ------ATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASSTTGY 423
Db    428 AAAQFNASPVAQSYLRNFLAAPPPQRAAMAAQL---QAVPGAAQYI-GLVESVAGSCNNY 483

Qy    424 VADQTVKLAKTVKDMGGEAITHTGASLR 451
Db    484 E-------LMTINYQFGDVDAH-GAMIR 503
```

RESULT 7
AAU74597
ID AAU74597 standard; Protein; 600 AA.
XX
AC AAU74597;
XX
DT 08-MAY-2002 (first entry)
XX
DE Antigenic fusion protein TbH9-DPV-MTI (Mtb61f).
XX
KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
KW tuberculostatic; immunogen; vaccine; TbH9-DPV-MTI; Mtb61f.
XX
OS Chimeric - Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 301
FT /note= "Encoded by GC"
XX
PN US2002009459-A1.
XX
PD 24-JAN-2002.
XX
PF 07-APR-1999; 99US-0287849.
XX
PR 13-MAR-1997; 97US-0818112.
PR 01-OCT-1997; 97US-0942578.
PR 18-FEB-1998; 98US-0025197.
PR 07-APR-1998; 98US-0056556.
PR 30-DEC-1998; 98US-0223040.
XX
PA (REED/) REED S G.
PA (SKEI/) SKEIKY Y A.
PA (DILL/) DILLON D C.
PA (ALDE/) ALDERSON M.
PA (CAMP/) CAMPOS-NETO A.
XX
PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
XX
DR WPI; 2002-171134/22.
DR N-PSDB; ABK14137.
XX
PT New fusion proteins of Mycobacterium tuberculosis antigens, useful for
PT diagnosing, treating or preventing M. tuberculosis infection,
PT particularly as vaccine for treating or preventing tuberculosis -
XX
PS Claim 1; Fig 10; 62pp; English.
XX
CC The invention relates to a purified polypeptide which induces an immune
CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
CC useful for diagnosing, treating or preventing M. tuberculosis infection,
CC particularly tuberculosis infection. In particular, the polypeptides are
CC useful as a vaccine formulation with an adjuvant to afford long-term
CC protection in animals against the development of tuberculosis. The
CC protein coding sequence may be used to encode a protein product for use
CC as an immunogen to induce and/or enhance an immune response to M.
CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
CC of the invention.
XX
SQ Sequence 600 AA;

Query Match 5.9%; Score 143; DB 23; Length 600;
Best Local Similarity 23.0%; Pred. No. 0.0029;
Matches 117; Conservative 58; Mismatches 165; Indels 168; Gaps 25;

```
Qy     20 TASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQA---------- 69
Db     88 TAGQAELTAAQVRV----------AAAAYETAYGLT---VPPPVIAENRAELMILIATNL 134

Qy     70 ------AVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL--- 120
Db    135 LGQNTPAIAVNEAEYGEMWAQDAAAMFGYAAATATATATLLPFEEAPEM--TSAGGLLEQ 192

Qy    121 -----RATPFAMAS-LLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVG------T 168
Db    193 AAAVEEASDTAAANQLMNNVPQALQQ---LAQPTQGTTP--SSKLGGLWKTVSPHRSPIS 247

Qy    169 KMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLA 228
Db    248 NMVSMANNHMS-MTNSGVSMTNTLSSMLKGFAPAAAAQ------AVQT-AAQNGVR---- 295

Qy    229 PALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQL 288
Db    296 -AMSSLGSSLGSSGLG----GGVAANLG---RAASVGS---------------------L 326

Qy    289 SEENDWLEAYKAIKSAS--------YSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASS 340
Db    327 SVPQAWAAANQAVTPAARALPLTSLTSAAERGPGQMLGGLPVG----------------- 369

Qy    341 LTQNGLALAGGFAGV----------------GKL--------------QEMATKNITDP- 369
Db    370 --QMGARAGGGLSGVLRVPPRPYVMPHSPAAGKLDPVDAVINTTCNYGQVVAALNATDPG 427

Qy    370 ------ATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASSTTGY 423
Db    428 AAAQFNASPVAQSYLRNFLAAPPPQRAAMAAQL---QAVPGAAQYI-GLVESVAGSCNNY 483

Qy    424 VADQTVKLAKTVKDMGGEAITHTGASLR 451
Db    484 E-------LMTINYQFGDVDAH-GAMIR 503
```

RESULT 8
AAM78810
ID AAM78810 standard; Protein; 1134 AA.
XX
AC AAM78810;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1472.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK51943.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 3747-3749; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1134 AA;

Query Match 5.9%; Score 142.5; DB 22; Length 1134;
Best Local Similarity 23.4%; Pred. No. 0.0077;
Matches 122; Conservative 70; Mismatches 248; Indels 81; Gaps 22;

```
Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKL--PPADSADGQAAVDVHNA 76
           |     ||||: | :|:::     : |: : ||  |      |  | | :| |:||   |
Db     267 RDIEQASLAAVS-QSLATRDDISVEALQEQLTSVVQEIGHLIDPIATAARGEAAQLGH-- 323

Qy      77 QITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQP 136
           ::| |          :||   |  :|     :: ||     |   |  |      | || |
Db     324 KVTQLAS------YFE---PLILAAVGVASKILDHQQQMT--VLDQTKTLAESALQ-MLY 371

Qy     137 AINKGDWLPAPLKPLTPLISGA--LSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAA 194
           |  :|   |        :   |  :  |:| :    :: |  ::  :     | : :||:
Db     372 AAKEGGGNPKAQHTHDAITEAAQLMKEAVDDI-MVTLNEAASEVGLVGGMVDAIAEAMSK 430

Qy     195 SVKRHSPSLARQVLDTGVAVQTYSARNAV-------RTVLAPALASRPAVQGAVDLG-VS 246
             :   |      :|    |  ||   ||        ::|  |       | |   | | ::
Db     431 LDEGTPPEPKGTFVDYQTTVVKYSKSIAVTAQEMMTKSVTNPEELGGLASQMTSDYGHLA 490

Qy     247 MAGGLAA------NAGFGNRLLSVQSRDH------QRGGALVLGLKDKEPKAQLSE-END 293
           | :||         ||  |   ||   |       |: ||| :   |   | :| |
Db     491 FQGQMAAATAEPEEIGFQIR-TRVQDLGHGCIFLVQKAGALQVCPTDSYTKRELIECARA 549

Qy     294 WLEAYKAIKSASYSG-----AALNAGKRMAGLPLDMATDAMGAVRSLVSASSL------T 342
           |     : ||  :|      | : |   ::|:  |: |  | |     ::| :
Db     550 VTEKVSLVLSALQAGNKGTQACITAATAVSGIIADLDTTIMFATAGTLNAENSETFADHR 609

Qy     343 QNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDP-- 400
           :| |  |        ||       :  |   :|| |    :   | |       :  : ||
Db     610 ENILKTAKALVEDTKLLVSGAASTPDKLAQAAQSSAATITQLAEVVKLGAASLGSDDPET 669

Qy     401 ------AVKKAESFIQDTVKSTASSTTGYVADQTV---------------KLAKTVKDMG 439
                |:|     : | : :|  : :  | | ::                 | |||| :
Db     670 QVVLINAIKDVAKALSDLISATKGAASKPVDDPSMYQLKGAAKVMVTNVTSLLKTVKAVE 729
```

Qy 440 GEAITHTGASLRNTVNNLRQRPA--READIEEGGTAASPSE 478
Db 730 DEATRGTRA-LEATIECIKQELTVFQSKDVPE--KTSSPEE 767

RESULT 9
AAM39293
ID AAM39293 standard; Protein; 1144 AA.
XX
AC AAM39293;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2438.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58449.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2438; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1144 AA;

Query Match 5.9%; Score 142.5; DB 22; Length 1144;

Best Local Similarity 23.4%; Pred. No. 0.0078;
Matches 122; Conservative 70; Mismatches 248; Indels 81; Gaps 22;

```
Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKL--PPADSADGQAAVDVHNA 76
           |    ||||: | :|:::     : |: : ||  |      |  | | :| |:||   |
Db     277 RDIEQASLAAVS-QSLATRDDISVEALQEQLTSVVQEIGHLIDPIATAARGEAAQLGH-- 333

Qy      77 QITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQP 136
           ::| |        :||   |  :|     :: ||      |   |  |      | || |
Db     334 KVTQLAS------YFE---PLILAAVGVASKILDHQQQMT--VLDQTKTLAESALQ-MLY 381

Qy     137 AINKGDWLPAPLKPLTPLISGA--LSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAA 194
           |  :|   |        :   |  :  |:| :    :: |  ::  :    | : :||:
Db     382 AAKEGGGNPKAQHTHDAITEAAQLMKEAVDDI-MVTLNEAASEVGLVGGMVDAIAEAMSK 440

Qy     195 SVKRHSPSLARQVLDTGVAVQTYSARNAV-------RTVLAPALASRPAVQGAVDLG-VS 246
             :   |     :|    | ||   ||       ::|  |      | |   | | ::
Db     441 LDEGTPPEPKGTFVDYQTTVVKYSKSIAVTAQEMMTKSVTNPEELGGLASQMTSDYGHLA 500

Qy     247 MAGGLAA------NAGFGNRLLSVQSRDH------QRGGALVLGLKDKEPKAQLSE-END 293
           | :||         ||  |   ||   |      |: ||| :   |   | :| |
Db     501 FQGQMAAATAEPEEIGFQIR-TRVQDLGHGCIFLVQKAGALQVCPTDSYTKRELIECARA 559

Qy     294 WLEAYKAIKSASYSG-----AALNAGKRMAGLPLDMATDAMGAVRSLVSASSL------T 342
           |    : ||  :|     | : |   ::|:  |: |  | |    ::| :
Db     560 VTEKVSLVLSALQAGNKGTQACITAATAVSGIIADLDTTIMFATAGTLNAENSETFADHR 619

Qy     343 QNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDP-- 400
           :| | |        ||      :  |   :|| |    :   | |      :  : ||
Db     620 ENILKTAKALVEDTKLLVSGAASTPDKLAQAAQSSAATITQLAEVVKLGAASLGSDDPET 679

Qy     401 ------AVKKAESFIQDTVKSTASSTTGYVADQTV--------------KLAKTVKDMG 439
                |:|     : | : :|  : :  | | ::               | |||| :
Db     680 QVVLINAIKDVAKALSDLISATKGAASKPVDDPSMYQLKGAAKVMVTNVTSLLKTVKAVE 739

Qy     440 GEAITHTGASLRNTVNNLRQRPA--READIEEGGTAASPSE 478
           ||   | | |  |:  ::|     :  |: |    :|| |
Db     740 DEATRGTRA-LEATIECIKQELTVFQSKDVPE--KTSSPEE 777
```

RESULT 10
AAM39312
ID AAM39312 standard; Protein; 1165 AA.
XX
AC AAM39312;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2457.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58468.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2457; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1165 AA;

Query Match 5.9%; Score 142.5; DB 22; Length 1165;
Best Local Similarity 23.4%; Pred. No. 0.008;
Matches 122; Conservative 70; Mismatches 248; Indels 81; Gaps 22;

```
Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKL--PPADSADGQAAVDVHNA 76
           |    ||||: | :|:::     : |: : ||  |      |  | | :| |:||   |
Db     298 RDIEQASLAAVS-QSLATRDDISVEALQEQLTSVVQEIGHLIDPIATAARGEAAQLGH-- 354

Qy      77 QITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQP 136
           ::| |        :||   |  :|     :: ||      |   |  |      | || |
Db     355 KVTQLAS------YFE---PLILAAVGVASKILDHQQQMT--VLDQTKTLAESALQ-MLY 402

Qy     137 AINKGDWLPAPLKPLTPLISGA--LSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAA 194
           |  :|   |        :   |  :  |:| :    :: |  ::  :    | : :||:
Db     403 AAKEGGGNPKAQHTHDAITEAAQLMKEAVDDI-MVTLNEAASEVGLVGGMVDAIAEAMSK 461

Qy     195 SVKRHSPSLARQVLDTGVAVQTYSARNAV-------RTVLAPALASRPAVQGAVDLG-VS 246
             :   |     :|    | ||   ||       ::|  |      | |   | | ::
Db     462 LDEGTPPEPKGTFVDYQTTVVKYSKSIAVTAQEMMTKSVTNPEELGGLASQMTSDYGHLA 521

Qy     247 MAGGLAA------NAGFGNRLLSVQSRDH------QRGGALVLGLKDKEPKAQLSE-END 293
           | :||         ||  |   ||   |      |: ||| :   |   | :| |
Db     522 FQGQMAAATAEPEEIGFQIR-TRVQDLGHGCIFLVQKAGALQVCPTDSYTKRELIECARA 580

Qy     294 WLEAYKAIKSASYSG-----AALNAGKRMAGLPLDMATDAMGAVRSLVSASSL------T 342
           |    : ||  :|     | : |   ::|:  |: |  | |    ::| :
Db     581 VTEKVSLVLSALQAGNKGTQACITAATAVSGIIADLDTTIMFATAGTLNAENSETFADHR 640

Qy     343 QNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDP-- 400
           :| | |        ||      :  |   :|| |    :   | |      :  : ||
Db     641 ENILKTAKALVEDTKLLVSGAASTPDKLAQAAQSSAATITQLAEVVKLGAASLGSDDPET 700

Qy     401 ------AVKKAESFIQDTVKSTASSTTGYVADQTV--------------KLAKTVKDMG 439
                |:|     : | : :|  : :  | | ::               | |||| :
Db     701 QVVLINAIKDVAKALSDLISATKGAASKPVDDPSMYQLKGAAKVMVTNVTSLLKTVKAVE 760

Qy     440 GEAITHTGASLRNTVNNLRQRPA--READIEEGGTAASPSE 478
           ||   | | |  |:  ::|     :  |: |    :|| |
Db     761 DEATRGTRA-LEATIECIKQELTVFQSKDVPE--KTSSPEE 798
```

RESULT 11
AAM79794
ID AAM79794 standard; Protein; 1177 AA.
XX
AC AAM79794;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3440.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK52927.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 336-337; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 1177 AA;

Query Match 5.8%; Score 139.5; DB 22; Length 1177;
Best Local Similarity 23.0%; Pred. No. 0.014;
Matches 120; Conservative 69; Mismatches 250; Indels 83; Gaps 21;

```
Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKL--PPADSADGQAAVDVH-N 75
           |    ||||: | :|:::     : |: : ||  |     |  | | :| |:||   |
Db     310 RDIEQASLAAVS-QSLATRDDISVEALQEQLTSVVQEIGHLIDPIATAARGEAAQLGHKG 368

Qy      76 AQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQ 135
           |: :  |             |  :|     :: ||      |   |  |      | || |
Db     369 TQLASYFE-----------PLILAAVGVASKILDHQQQMT--VLDQTKTLAESALQ-ML 413

Qy     136 PAINKGDWLPAPLKPLTPLISGA--LSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMA 193
           |  :|  |        :   |  :  |:| :    :: |  ::  :    | : :||:
Db     414 YAAKEGGGNPKAQHTHDAITEAAQLMKEAVDDI-MVTLNEAASEVGLVGGMVDAIAEAMS 472

Qy     194 ASVKRHSPSLARQVLDTGVAVQTYSARNAV-------RTVLAPALASRPAVQGAVDLG-V 245
             :  |      :|     |  ||   ||        ::|  |       | |   | | :
Db     473 KLDEGTPPEPKGTFVDYQTTVVKYSKAIAVTAQEMMTKSVTNPEELGGLASQMTSDYGHL 532

Qy     246 SMAGGLAA------NAGFGNRLLSVQSRDH------QRGGALVLGLKDKEPKAQLSE-EN 292
           :  | :||        ||  |   ||   |      |: ||| :   |   | :| |
Db     533 AFQGQMAAATAEPEEIGFQIR-TRVQDLGHGCIFLVQKAGALQVCPTDSYTKRELIECAR 591

Qy     293 DWLEAYKAIKSASYSG-----AALNAGKRMAGLPLDMATDAMGAVRSLVSASSL------ 341
             |    : ||  :|      | : |   ::|:  |: |  | |     ::| :
Db     592 AVTEKVSLVLSALQAGNKGTQACITAATAVSGIIADLDTTIMFATAGTLNAENSETFADH 651

Qy     342 TQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDP- 400
           :| | |        ||      :  |   :|| |    :   | |       :  : ||
Db     652 RENILKTAKALVEDTKLLVSGAASTPDKLAQAAQSSAATITQLAEVVKLGAASLGSDDPE 711

Qy     401 -------AVKKAESFIQDTVKSTASSTTGYVADQTV---------------KLAKTVKDM 438
                  |:|      : | : :|  : :  | | ::               | |||| :
Db     712 TQVVLINAIKDVAKALSDLISATKGAASKPVDDPSMYQLKGAAKVMVTNVTSLLKTVKAV 771

Qy     439 GGEAITHTGASLRNTVNNLRQRPA--READIEEGGTAASPSE 478
             ||   | | |  |:  ::|     :  |: |    :|| |
Db     772 EDEATRGTRA-LEATIECIKQELTVFQSKDVPE--KTSSPEE 810
```

RESULT 12
AAM41079
ID AAM41079 standard; Protein; 1177 AA.
XX
AC AAM41079;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 6010.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX

DR WPI; 2001-442253/47.
DR N-PSDB; AAI60235.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 6010; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1177 AA;

Query Match 5.8%; Score 139.5; DB 22; Length 1177;
Best Local Similarity 23.0%; Pred. No. 0.014;
Matches 120; Conservative 69; Mismatches 250; Indels 83; Gaps 21;

```
Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKL--PPADSADGQAAVDVH-N 75
           |    ||||: | :|:::     : |: : ||  |      |  | | :| |:||   |
Db     310 RDIEQASLAAVS-QSLATRDDISVEALQEQLTSVVQEIGHLIDPIATAARGEAAQLGHKG 368

Qy      76 AQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQ 135
             |: :  |             |  :|     :: ||      |   | |     | || |
Db     369 TQLASYFE-----------PLILAAVGVASKILDHQQQMT--VLDQTKTLAESALQ-ML 413

Qy     136 PAINKGDWLPAPLKPLTPLISGA--LSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMA 193
             |  :|   |          :   |  :  |:| :    :: |  ::  :     | : :||:
Db     414 YAAKEGGGNPKAQHTHDAITEAAQLMKEAVDDI-MVTLNEAASEVGLVGGMVDAIAEAMS 472

Qy     194 ASVKRHSPSLARQVLDTGVAVQTYSARNAV-------RTVLAPALASRPAVQGAVDLG-V 245
              :   |      :|     |  ||   ||       ::|  |       | |   | | :
Db     473 KLDEGTPPEPKGTFVDYQTTVVKYSKAIAVTAQEMMTKSVTNPEELGGLASQMTSDYGHL 532

Qy     246 SMAGGLAA------NAGFGNRLLSVQSRDH------QRGGALVLGLKDKEPKAQLSE-EN 292
             :  | :||        ||  |   ||   |      |: ||| :   |   | :| |
Db     533 AFQGQMAAATAEPEEIGFQIR-TRVQDLGHGCIFLVQKAGALQVCPTDSYTKRELIECAR 591

Qy     293 DWLEAYKAIKSASYSG-----AALNAGKRMAGLPLDMATDAMGAVRSLVSASSL------ 341
             |    : ||  :|      | : |   ::|:  |: |  | |    ::| :
Db     592 AVTEKVSLVLSALQAGNKGTQACITAATAVSGIIADLDTTIMFATAGTLNAENSETFADH 651

Qy     342 TQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDP- 400
            :| |  |         ||      :  |   :|| |    :   | |       :  : ||
Db     652 RENILKTAKALVEDTKLLVSGAASTPDKLAQAAQSSAATITQLAEVVKLGAASLGSDDPE 711

Qy     401 -------AVKKAESFIQDTVKSTASSTTGYVADQTV--------------KLAKTVKDM 438
                  |:|     : | : :|  : :   | | ::              | |||| :
Db     712 TQVVLINAIKDVAKALSDLISATKGAASKPVDDPSMYQLKGAAKVMVTNVTSLLKTVKAV 771

Qy     439 GGEAITHTGASLRNTVNNLRQRPA--READIEEGGTAASPSE 478
             ||   | | |  |:  ::|     :  |: |    :|| |
Db     772 EDEATRGTRA-LEATIECIKQELTVFQSKDVPE--KTSSPEE 810
```

RESULT 13
AAM41098
ID AAM41098 standard; Protein; 1177 AA.
XX
AC AAM41098;

XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 6029.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI60254.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 6029; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1177 AA;

Query Match 5.8%; Score 139.5; DB 22; Length 1177;
Best Local Similarity 23.0%; Pred. No. 0.014;
Matches 120; Conservative 69; Mismatches 250; Indels 83; Gaps 21;

```
Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKL--PPADSADGQAAVDVH-N 75
           |    ||||: | :|:::     : |: : ||  |      |  | | :| |:||   |
Db     310 RDIEQASLAAVS-QSLATRDDISVEALQEQLTSVVQEIGHLIDPIATAARGEAAQLGHKG 368

Qy      76 AQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQ 135
             |: :  |             |  :|     :: ||      |   | |     | || |
Db     369 TQLASYFE-----------PLILAAVGVASKILDHQQQMT--VLDQTKTLAESALQ-ML 413
```

```
Qy     136 PAINKGDWLPAPLKPLTPLISGA--LSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMA 193
Db     414 YAAKEGGGNPKAQHTHDAITEAAQLMKEAVDDI-MVTLNEAASEVGLVGGMVDAIAEAMS 472

Qy     194 ASVKRHSPSLARQVLDTGVAVQTYSARNAV-------RTVLAPALASRPAVQGAVDLG-V 245
Db     473 KLDEGTPPEPKGTFVDYQTTVVKYSKAIAVTAQEMMTKSVTNPEELGGLASQMTSDYGHL 532

Qy     246 SMAGGLAA------NAGFGNRLLSVQSRDH------QRGGALVLGLKDKEPKAQLSE-EN 292
Db     533 AFQGQMAAATAEPEEIGFQIR-TRVQDLGHGCIFLVQKAGALQVCPTDSYTKRELIECAR 591

Qy     293 DWLEAYKAIKSASYSG-----AALNAGKRMAGLPLDMATDAMGAVRSLVSASSL------ 341
Db     592 AVTEKVSLVLSALQAGNKGTQACITAATAVSGIIADLDTTIMFATAGTLNAENSETFADH 651

Qy     342 TQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDP- 400
Db     652 RENILKTAKALVEDTKLLVSGAASTPDKLAQAAQSSAATITQLAEVVKLGAASLGSDDPE 711

Qy     401 -------AVKKAESFIQDTVKSTASSTTGYVADQTV---------------KLAKTVKDM 438
Db     712 TQVVLINAIKDVAKALSDLISATKGAASKPVDDPSMYQLKGAAKVMVTNVTSLLKTVKAV 771

Qy     439 GGEAITHTGASLRNTVNNLRQRPA--READIEEGGTAASPSE 478
Db     772 EDEATRGTRA-LEATIECIKQELTVFQSKDVPE--KTSSPEE 810
```

RESULT 14
AAY75477
ID AAY75477 standard; Protein; 723 AA.
XX
AC AAY75477;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 719 protein sequence SEQ ID NO:2428.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ54239.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 1164; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 723 AA;

Query Match 5.6%; Score 135; DB 21; Length 723;
Best Local Similarity 22.4%; Pred. No. 0.017;
Matches 119; Conservative 76; Mismatches 224; Indels 112; Gaps 25;

```
Qy      14 ATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDV 73
Db     166 ATEGAQQIKD--LALELVEKNGGTHDKALDLISGMMTTGLNFAQTKNEAQAAYAFALASE 223

Qy      74 HNAQITA-LIET--------RASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATP 124
Db     224 GSGEDTAKLIKTLKDGGMSGKDLQLGLEHVLQSGLDGTFEVRDMVRELPSLLSAAQQAGM 283

Qy     125 FAMAS---LLQYMQPAINKGDWLPAPL---------KPLTPLISGAL---------SGAM 163
Db     284 NGVGGLDLYLLSLLQSAANKSG-SPAEAATNVQNLLSKTLSPDTIGRLKKMANPNDPKKGV 342

Qy     164 DQVGTKMMDRATGD--LHYLSASPDRLHDAMAASVKRHSPSLARQVL-DTGVAVQTYSAR 220
Db     343 DWIGSVVQGKQNGENAVQVLS----RLADAMLVKDKQYQDYKKRAAAGDKTAAEQANMLK 398

Qy     221 NAVRTVLAPALASRPAVQGAVDL----------GVSMAGGLAANAGFGNRLLSVQSRDH 269
Db     399 GALLAQLLPDLQAKQGLLAATDMTQIREYMASLAGVTLDNGKIAKNNEA-RMLSAAAQQE 457

Qy     270 QRGGALVLGLKDKEPKAQLSEE-----NDWLEAYKAIKSASYSGAALNAGKRMAGLPLDM 324
Db     458 Q----------QESLAMLRESLTGTLVDMETSFKKL-AAEYPNATL------------ 492

Qy     325 ATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQEMATKNI-TDPATKAAVSQLTNLAG 383
Db     493 ---ALQALTTAATAASAAM--LLTAGGGKGAGFLKDVGSKALGWGKASAGGVAAGATAAG 547

Qy     384 SAAVFAGWTTAA---LTTDPAVKKAESFIQDTVKSTASSTTGYVADQTVKLAKTVKDMGG 440
Db     548 GKLL--SWGKSAGSGLMNNPALVKRAGLLGMLLYSES------LGDGTLP-----KGLRG 594

Qy     441 EAITHTGASLRNTV--NNLRQRPAREADIEEGGT---AASPSEIPFRPMRS 486
Db     595 ---TKTTPEMINRLKNNGIRFEPAPKREQARGGVPQYLAAPSAQPTDKMLS 642
```

RESULT 15
AAE17572
ID AAE17572 standard; Protein; 729 AA.
XX
AC AAE17572;
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species MTB72F fusion protein.
XX
KW Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB72F; Ra12-TbH9-Ra35 protein.
XX

OS Mycobacterium sp.
XX
FH Key Location/Qualifiers
FT Region 8..139
FT /label= Ra12_protein
FT Region 142..532
FT /label= TbH9FL_protein
FT Region 535..729
FT /label= Ra35_protein
XX
PN WO200198460-A2.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-US19959.
XX
PR 20-JUN-2000; 2000US-0597796.
PR 01-FEB-2001; 2001US-265737P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Alderson M;
XX
DR WPI; 2002-147798/19.
DR N-PSDB; AAD28342.
XX
PT Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a
PT subject -
XX
PS Claim 6; Fig 5; 136pp; English.
XX
CC The present invention relates to fusion proteins containing at least
CC two Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected
CC with Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is a DNA encoding
CC Mycobacterium species MTB72F (Ra12-TbH9-Ra35) fusion protein.
XX
SQ Sequence 729 AA;

Query Match 5.5%; Score 133.5; DB 23; Length 729;
Best Local Similarity 21.0%; Pred. No. 0.023;
Matches 116; Conservative 70; Mismatches 178; Indels 189; Gaps 25;

```
Qy     20 TASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQA---------- 69
          ||  | | :: ||           | | | | :      :||   |: :|
Db    221 TAGQAELTAAQVRV----------AAAAYETAYGLT---VPPPVIAENRAELMILIATNL 267

Qy     70 ------AVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL--- 120
               |: |: |:   :    |: :       ||  .|    |:   :  |::| |
Db    268 LGQNTPAIAVNEAEYGEMWAQDAAAMFGYAAATATATATLLPFEEAPEM--TSAGGLLEQ 325

Qy    121 -----RATPFAMAS-LLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVG------T 168
               |:  | |: |:  :  |: :   |  | : || |  | |     |       :
Db    326 AAAVEEASDTAAANQLMNNVPQALQQ---LAQPTQGTTP--SSKLGGLWKTVSPHRSPIS 380

Qy    169 KMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLA 228

          |:  |   :  :: |   : : ::: :|  :|: |||      |||| :|:| ||
Db    381 NMVSMANNHMS-MTNSGVSMTNTLSSMLKGFAPAAARQ------AVQT-AAQNGVR---- 428

Qy    229 PALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQL 288
          |::|  :  |:  ||    ||:||| |   | || |                      |
Db    429 -AMSSLGSSLGSSGLG----GGVAANLG---RAASVGS---------------------L 459

Qy    289 SEENDWLEAYKAIKSAS--------YSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASS 340
          |     |  | :|:  |:          | |     |: : |||:
Db    460 SVPQAWAAANQAVTPAARALPLTSLTSAAERGPGQMLGGLPVG---------------- 502

Qy    341 LTQNGLALAGGFAGVGKL--------QEMATKNITDPA---------------TKAAVSQ 377
            | |     || :|| ::           |  :|  ||               | |:|
Db    503 --QMGARAGGGLSGVLRVPPRPYVMPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQ 560

Qy    378 L--------TNLAGSAAVFAGWTTAALTTDP-----------AVKKAESFIQDTVKSTA 417
          :        | |  : || ||      :  ||                  :|   : ::
Db    561 VGPQVVNINTKLGYNNAVGAG---TGIVIDPNGVVLTNNHVIAGATDINAFSVGSGQTYG 617

Qy    418 SSTTGYVADQTVKLAK--------TVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEE 469
             ||    | | : :        :    || |:     :: |:
Db    618 VDVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMGNS---------------- 661

Qy    470 GGTAASPSEIPFR 482
          ||   :|  :| |
Db    662 GGQGGTPRAVPGR 674
```

Search completed: January 2, 2003, 15:18:52
Job time : 40.4645 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:14:20 ; Search time 9.49024 Seconds
(without alignments)
2124.025 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHINRRVQQPPVTATDSFRT.....IEEGGTAASPSEIPFRPMRS 486

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	155.5	6.4	1120	1	STFR_ECOLI	P76072 escherichia
2	136	5.6	505	1	FLJB_SALTY	P52616 salmonella
3	131.5	5.5	3591	1	FHAB_BORPE	P12255 bordetella
4	131	5.4	556	1	FTHS_CLOAC	P13419 clostridium
5	129	5.3	676	1	MUC1_MESAU	Q60528 mesocricetu
6	129	5.3	732	1	YAGR_ECO57	Q8x6j4 escherichia
7	128.5	5.3	1025	1	SLAP_CAUCR	P35828 caulobacter
8	127.5	5.3	401	1	YOPB_YERPS	Q06114 yersinia ps
9	127	5.3	1068	1	HIPR_HUMAN	O75146 homo sapien
10	127	5.3	1238	1	SBCC_RHOCA	O68032 rhodobacter
11	126.5	5.2	397	1	NCCB_ALCXX	Q44585 alcaligenes
12	125.5	5.2	492	1	FLIC_SALRU	P06175 salmonella
13	125.5	5.2	1065	1	SED4_YEAST	P25365 saccharomyc
14	125.5	5.2	1068	1	HIPR_MOUSE	Q9jky5 mus musculu
15	125.5	5.2	2541	1	TALI_MOUSE	P26039 mus musculu
16	124.5	5.2	2541	1	TALI_HUMAN	Q9y490 homo sapien
17	122.5	5.1	452	1	PPOX_MYCTU	O53230 mycobacteri
18	122.5	5.1	500	1	FLJB_SALAE	P52615 salmonella
19	122	5.1	756	1	Y4SI_RHISN	P55652 rhizobium s
20	121.5	5.0	488	1	FLIC_SALCH	P06176 salmonella
21	121	5.0	1140	1	YM96_YEAST	Q04893 saccharomyc
22	120.5	5.0	739	1	OCT1_CHICK	P15143 gallus gall
23	119	4.9	732	1	YAGR_ECOLI	P77489 escherichia
24	118.5	4.9	1714	1	SYEP_DROME	P28668 drosophila
25	118	4.9	1218	1	MGPC_MYCPN	Q50341 mycoplasma
26	117	4.9	570	1	FLIF_RHOSH	Q53151 rhodobacter
27	117	4.9	583	1	YP65_MYCTU	Q50733 mycobacteri
28	117	4.9	1508	1	BCSC_XANAC	P58938 xanthomonas
29	116	4.8	774	1	STF_LAMBD	P03764 bacteriopha
30	116	4.8	934	1	CAPP_MYCLE	P46710 mycobacteri
31	116	4.8	1306	1	MSB2_YEAST	P32334 saccharomyc
32	115.5	4.8	397	1	DP3B_MYCSM	P52851 mycobacteri
33	115.5	4.8	2249	1	OMPA_RICRI	P15921 rickettsia
34	115	4.8	342	1	GCP_HAEIN	P43764 haemophilus
35	115	4.8	559	1	FTHS_MOOTH	P21164 moorella th
36	115	4.8	802	1	BCB1_ACEXY	P37716 acetobacter
37	115	4.8	1756	1	TRI1_ECOLI	P14565 escherichia
38	114.5	4.7	439	1	Y579_CHLTR	O84583 chlamydia t
39	114.5	4.7	743	1	OCT1_HUMAN	P14859 homo sapien
40	114.5	4.7	920	1	MML7_MYCTU	P96289 mycobacteri
41	114.5	4.7	1783	1	RAA3_CHLRE	Q9fec4 chlamydomon
42	114.5	4.7	2090	1	N214_HUMAN	P35658 homo sapien
43	114	4.7	575	1	FLA2_CAMJE	P22251 campylobact
44	114	4.7	1325	1	BCC3_ACEXY	Q9wx63 acetobacter
45	114	4.7	1756	1	TRI2_ECOLI	P22706 escherichia

ALIGNMENTS

RESULT 1
STFR_ECOLI
ID STFR_ECOLI STANDARD; PRT; 1120 AA.
AC P76072; P77560;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Side tail fiber protein homolog from lambdoid prophage Rac.
GN STFR OR B1372.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -!- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AE000234; AAC74454.1; ALT_INIT.
DR EMBL; D90774; BAA14966.1; -.
DR EMBL; D90775; BAA14975.1; -.
DR EcoGene; EG13370; stfR.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR005003; Phage_fiber.
DR InterPro; IPR005068; Phage_fiber_2.
DR Pfam; PF03335; Phage_fiber; 6.
DR Pfam; PF03406; Phage_fiber_2; 1.
KW Hypothetical protein; Fiber protein; Repeat; Complete proteome.

```
SQ   SEQUENCE   1120 AA;  113779 MW;  542E59D71EE795B4 CRC64;

  Query Match             6.4%;  Score 155.5;  DB 1;  Length 1120;
  Best Local Similarity   23.7%;  Pred. No. 0.086;
  Matches  121;  Conservative   60;  Mismatches  220;  Indels  109;  Gaps   21;

Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQI 78
           | ||  :  ::: :  :||       | :  |   ||     | :: ||||    :|
Db     115 RNASAVAQNTAAAKKSASDASTSAREAATHAAD---AADSARAASTSAGQAASSAQSASS 171

Qy      79 TALIETRASRLHFEGETPATIAD-TFAKAEKLDRLATTTSGALRATPFAMASLLQYMQPA 137
           :|              | || |  : | ||     | |::|| : :    :: ||    :
Db     172 SA----------GTASTKATEASKSAAAAESSKSAAATSAGAAKTSETNASASLQSAATS 221

Qy     138 INKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAA--S 195
           :           |   | | : | |   :|   ::: : :  |::      | ||  |
Db     222 ASTA----------TTKASEAATSARDAAASKEAAKSS-ETNASSSASSAASSATAAGNS 270

Qy     196 VK---------RHSPSLARQVLDTGVAVQT--YSARNAVRTVLAPALASRPAVQGAVDLG 244
           |         | | : | |        :|   |: :|  |    | ||  |   : :
Db     271 AKAAKTSETNARSSETAAGQSASAAAGSKTAAASSASAASTSAGQASASATAAGKSAESA 330

Qy     245 VSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSA 304
           | |       ||      |  :|          |  |  |: || :   |: |   ::
Db     331 ASSASTATTKAGEATEQASAAARS-------ASAAKTSETNAKASETS--AESSKTAAAS 381

Qy     305 SYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQEMATK 364
           | | || :|     |    | ||     | :|  ||:: :       ||
Db     382 SASSAASSASS--ASASKDEATRQASAAKS--SATTASTKATEAAGS------------- 424

Qy     365 NITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESF-----IQD--TVK--- 414
                || || |: |  | |||  |         : | |:||      ::|  | |
Db     425 -----ATAAAQSKST--AESAATRA---------ETAAKRAEDIASAVALEDASTTKKGI 468

Qy     415 ---STASSTTGYVADQTVKLAKTV---------KDMGGEAITHTGASLRNTVNNLR---- 458
               |:|:::|      | |  |:          ||  |  |   |  | | :| :
Db     469 VQLSSATNSTSETLAATPKAVKSAYDNAEKRLQKDQNGADIPDKGCFL-NNINAVSKTDF 527

Qy     459 --QRPAREADIEEGGTAASPSEIPFRPMRS 486
             :|  |   :     | |    |   |||
Db     528 ADKRGMRYVRVNAPAGATSGKYYPVVVMRS 557


RESULT 2
FLJB_SALTY
ID   FLJB_SALTY     STANDARD;      PRT;   505 AA.
AC   P52616; P97159;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phase-2 flagellin.
GN   FLJB OR H2 OR STM2771.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SL 375;
RX   MEDLINE=95325331; PubMed=7541401;
RA   Vanegas R.A., Joys T.M.;
RT   "Molecular analyses of the phase-2 antigen complex 1,2,. . of
RT   Salmonella spp.";
RL   J. Bacteriol. 177:3863-3864(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
RN   [3]
RP   SEQUENCE OF 1-37 FROM N.A.
RX   MEDLINE=82049491; PubMed=6271461;
RA   Silverman M., Zieg J., Mandel G., Simon M.;
RT   "Analysis of the functional components of the phase variation
RT   system.";
RL   Cold Spring Harb. Symp. Quant. Biol. 45:17-26(1981).
RN   [4]
RP   SEQUENCE OF 482-505 FROM N.A.
RC   STRAIN=SJ2353;
RA   Mingorance J., Tanaka S., Tominaga A., Enomoto M.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC       FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC   -!- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC       BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC       PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17177; AAC43354.1; -.
DR   EMBL; AE008826; AAL21657.1; -.
DR   EMBL; V01370; CAA24655.1; -.
DR   EMBL; D26168; BAA05156.1; -.
DR   StyGene; SG10564; fljB.
DR   InterPro; IPR001492; FlagellinN.
DR   InterPro; IPR001029; Flagellin_C.
DR   Pfam; PF00669; Flagellin_N; 1.
DR   Pfam; PF00700; Flagellin_C; 1.
DR   PRINTS; PR00207; FLAGELLIN.
DR   ProDom; PD000316; Flagellin_C; 1.
KW   Flagella; Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CONFLICT     37     37       I -> S (IN REF. 3).
SQ   SEQUENCE   505 AA;  52404 MW;  3A5CC404AF7AF88B CRC64;

  Query Match             5.6%;  Score 136;  DB 1;  Length 505;
  Best Local Similarity   22.4%;  Pred. No. 0.44;
  Matches  116;  Conservative   60;  Mismatches  179;  Indels  164;  Gaps   25;

Qy      63 DSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL-- 120
           | | |||  :   | |  |  |:|||   :| :               :| || |||
Db      42 DDAAGQAIANRFTANIKGL--TQASRNANDGIS---------------IAQTTEGALNE 83

Qy     121 ------RATPFAMASL--------LQYMQPAI----NKGDWLPAPLKPLTPLISGALSGA 162
                |    |: |          |  :|  |     |: | :    :      :|    |
Db      84 INNNLQRVRELAVQSANSTNSQSDLDSIQAEITQRLNEIDRVSGQTQ-----FNGVKVLA 138

Qy     163 MD-----QVGTKMMDRATGDLHYLSASP---DRLHDAMAASVKRHSPSLARQVLDTGVAV 214
           |      |||  :    ||  :::      | |:   |  ||            || |
Db     139 QDNTLTIQVGANDGETIDIDLKQINSQTLGLDSLNVQKAYDVK----------DTAVTT 187

Qy     215 QTYSARNAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFGNRLL----------SV 264
           : |:      |     |:  |  |  :  |: ||        |:           :
Db     188 KAYANNGTTLDVSGLDDAAIKAATGGTNGTASVTGGAVKFDADNNKYFVTIGGFTGADAA 247

Qy     265 QSRDHQ----RGGALVL---GLKDKEPK-AQLSEENDWLEAYKAIKSASYSGAALNAGKR 316
           :: |::      | : |     |  |  |    |:   |: ||    | :  |
Db     248 KNGDYEVNVATDGTVTLAAGATKTTMPAGATTKTEVQELKDTPAVVSADAKNALIAGG-- 305

Qy     317 MAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFA-------------GVGKLQEMA 362
               :| |||| ||   ||  |   :||  : ||:|                | ::
Db     306 -----VD-ATDANGA--ELVKMSYTDKNGKTIEGGYALKAGDKYYAADYDEATGAIKAKT 357
```

```
Qy    363 TK-NITDPATKAAVSQLTNLAGSAAVFA--GWT------------------TAALTTDP 400
          |      |  || | :||  : |    |  ||                  || ||:
Db    358 TSYTAADGTTKTAANQLGGVDGKTEVVTIDGKTYNASKAAGHDFKAQPELAEAAAKTTEN 417

Qy    401 AVKKAESFIQ--DTVKSTASSTTGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLR 458
           ::| :: :   | ::|   :              |::   |||: |   ||||||
Db    418 PLQKIDAALAQVDALRSDLGA--------------VQNRFNSAITNLG----NTVNNLS 458

Qy    459 QRPAREAD--------------IEEGGTA--ASPSEIP 480
          :  :|  |               ::: ||:  |  :::|
Db    459 EARSRIEDSDYATEVSNMSRAQILQQAGTSVLAQANQVP 497
```

RESULT 3
FHAB_BORPE
ID FHAB_BORPE STANDARD; PRT; 3591 AA.
AC P12255;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Filamentous hemagglutinin.
GN FHAB.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90355839; PubMed=2388559;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous
RT haemagglutinin: a protein processed from an unusually large
RT precursor.";
RL Mol. Microbiol. 4:787-800(1990).
RN [2]
RP SEQUENCE OF 1-3261 FROM N.A.
RX MEDLINE=89202384; PubMed=2539596;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
RT sequence and crucial role in adherence.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC -!- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
CC INFECTION.
CC -!- SUBCELLULAR LOCATION: SURFACE.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M60351; AAA22974.1; -.
DR EMBL; M60351; AAA22975.1; ALT_INIT.
DR EMBL; M60351; AAA22976.1; ALT_INIT.
KW Antigen; Hemagglutinin.
SQ SEQUENCE 3591 AA; 367420 MW; EF7418B30D6E5138 CRC64;

Query Match 5.5%; Score 131.5; DB 1; Length 3591;
Best Local Similarity 22.2%; Pred. No. 8.7;
Matches 111; Conservative 77; Mismatches 195; Indels 117; Gaps 24;

```
Qy     33 SVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQITALIETRA------ 86
          : || |   :||      | : |  ::      ||: || : :|   ||   ||
Db    414 AASSRQALSVNAGGALKADKLSATRRV----DVDGKQAVALGSASSNAL-SVRAGGALKA 468

Qy     87 ------SRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYM-QPAIN 139
                ||  :|:   |:    :       |: :  | |||     :: |:   |  :
Db    469 GKLSATGRLDVDGKQAVTLGSVASDG----ALSVSAGGNLRANELVSSAQLEVRGQREVA 524

Qy    140 KGDWLPAPLKPLTPLISGALS-------GAMDQVG-------------TKMMDRATGDL 178
```

```
          |   :  : :| : :|||:       ||:   |                ::  |   ||
Db    525 LDD--ASSARGMTVVAAGALAARNLQSKGAIGVQGGEAVSVANANSDAELRVRGRGQVDL 582

Qy    179 HYLSASPDRLHDAMAASVKRHS-PSLARQVLDTGVAVQTYSARN------------AVR 224
          | |||       |  | :      :: |   |: | |  : | :             |
Db    583 HDLSA-------ARGADISGEGRVNIGRARSDSDVKVSAHGALSIDSMTALGAIGVQAGG 635

Qy    225 TVLAPALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRG-GALVLGLKDKE 283
          :| |  : || |         |:::|| | | |       |||     |   |  : ::|
Db    636 SVSAKDMRSRGA--------VTVSGGGAVNLG------DVQSDGQVRATSAGAMTVRDVA 681

Qy    284 PKAQLS-EENDWLEA--YKAIKSASYSGA-ALNAGKRMAGLPLDMATDAMGAVRSLVSAS 339
            |:  :  | |:|   |:  : : :|  |:      ||  | :::|   |: || :
Db    682 AAADLALQAGDALQAGFLKSAGAMTVNGRDAVRLDGAHAGGQLRVSSDGQAALGSLAAKG 741

Qy    340 SLTQNGLALAGGFAGVGKLQEMATKNIT----------DPATKAAVSQLTNL------AG 383
            ||  :    |   | | :|: :   ::|          : |:: |:|    |      |
Db    742 ELTVS----AARAATVAELKSLDNISVTGGERVSVQSVNSASRVAISAHGALDVGKVSAK 797

Qy    384 SAAVFAGWTTA---ALTTDPAVKKA--ESFIQDTVKSTASSTTGYVADQT---VKLAKTV 435
          |     ||      :| :| |:  :  ::   |  :| |  : |     |   |: | ::
Db    798 SGIGLEGWGAVGADSLGSDGAISVSGRDAVRVDQARSLADISLGAEGGATLGAVEAAGSI 857

Qy    436 KDMGGEAIT----HTGASLR 451
            ||  :      |    :|
Db    858 DVRGGSTVAANSLHANRDVR 877
```

RESULT 4
FTHS_CLOAC
ID FTHS_CLOAC STANDARD; PRT; 556 AA.
AC P13419;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Formate--tetrahydrofolate ligase (EC 6.3.4.3) (Formyltetrahydrofolate
DE synthetase) (FHS) (FTHFS).
GN FHS.
OS Clostridium acidi-urici.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1556;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88257049; PubMed=2838464;
RA Whitehead T.R., Rabinowitz J.C.;
RT "Nucleotide sequence of the Clostridium acidiurici ('Clostridium
RT acidi-urici') gene for 10-formyltetrahydrofolate synthetase shows
RT extensive amino acid homology with the trifunctional enzyme
RT C1-tetrahydrofolate synthase from Saccharomyces cerevisiae.";
RL J. Bacteriol. 170:3255-3261(1988).
CC -!- CATALYTIC ACTIVITY: ATP + formate + tetrahydrofolate = ADP +
CC phosphate + 10-formyltetrahydrofolate.
CC -!- PATHWAY: NECESSARY FOR THE BIOSYNTHESIS OF PURINES, THYMYDYLATE,
CC METHIONINE, HISTIDINE, PANTOTHENATE, AND FORMYL TRNA-MET.
CC -!- SIMILARITY: TO FTHFS COMPONENT OF EUKARYOTIC MULTIFUNCTIONAL
CC ENZYMES.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M21507; AAA53187.1; -.
DR PIR; A28185; A28185.
DR InterPro; IPR000559; Fmtethyd_synth.
DR Pfam; PF01268; FTHFS; 1.
DR PROSITE; PS00721; FTHFS_1; 1.
DR PROSITE; PS00722; FTHFS_2; 1.
KW Ligase; One-carbon metabolism; ATP-binding.

```
FT   NP_BIND      65     72       ATP (BY SIMILARITY).
SQ   SEQUENCE   556 AA;  59588 MW;  72D9D485863F0334 CRC64;

  Query Match             5.4%;  Score 131;  DB 1;  Length 556;
  Best Local Similarity   22.3%;  Pred. No. 0.97;
  Matches  111;  Conservative   65;  Mismatches  160;  Indels  162;  Gaps   26;

Qy      29 SSVRSVSSDQQREINAIADYLTDH-----VFAAHKLPPA--------DSADGQAAVDVHN 75
           : ::     | : |  :|: :  |      ::  :|   :          |  ||:  :
Db       3 TDIQIAQEAQMKHIKDVAELIDIHEDDLELYGKYKAKVSLDVLDQLKDKPDGKLVL---- 58

Qy      76 AQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQ 135
             :||:  | |     ||:|   |  :      |::|   || |||             :
Db      59 --VTAINPTPAG----EGKTTTNIGLSMG----LNKLGKKTSTALR------------E 95

Qy     136 PAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMD---RATGDLHYLSASPDRLHDAM 192
           |:          | |   :  ||  |   ||   | |     ||| | ::::    |  :
Db      96 PS----------LGPSFGVKGGAAGGGYAQV-VPMADINLHFTGDFHAITSA----HSLL 140

Qy     193 AASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGAVDLGVSMAGGLA 252
           || |  |         |  | |::  : |   : |:   :  |   :  | ||   | |:
Db     141 AALVDNH--------LHHGNALRIDTNRIVWKRVV--DMNDRALRKIVVGLG-GKAQGIT 189

Qy     253 ANAGF-------------------------GNRLL-------SVQSRDHQRGGALVLGLK 280
           ||                            || ::       :|:::| :  ||| | ||
Db     190 REDGFDITVASEIMAILCLANDREDLKERLGNMVVAYNVDGDAVRAKDLEAQGALTLILK 249

Qy     281 DK-----------------EPKAQLSEE-NDWLEAYKAIKSASYS------GAALNAGK 315
           |                          | | ::   |  |    |:|:  |:      || | | |
Db     250 DAINPNIVQTLENTPAFIHGGPFANIAHGCNSVLATKLALKTGDYAVTEAGFGADLGAEK 309

Qy     316 ------RMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDP 369
                | |||  |:|          :|:    |   | : || |     ::: |:|:
Db     310 FFDIKCRYAGLNPDVAV--------IVA----TVRALKMHGGVA----KEDLGTENL--- 350

Qy     370 ATKAAVSQLTNL---AGSAAVFAGWTTAALTTDPAVKKAE-SFIQDTVKS-----TASST 420
                 |   :|||     : | |   :  |:   |   :|| : | |         |
Db     351 --DALAKGMTNLERHIENVAKFGVPSVVAINAFPTDTEAEKQLVFDKCKEMGVDVAISDV 408

Qy     421 TGYVADQTVKLAKTVKDM 438
                 |  |:||: | |:
Db     409 FAKGGDGGVELAQKVIDV 426


RESULT 5
MUC1_MESAU
ID   MUC1_MESAU     STANDARD;      PRT;   676 AA.
AC   Q60528;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Mucin 1 precursor.
GN   MUC1.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Tracheal epithelium;
RX   MEDLINE=96326118; PubMed=8703480;
RA   Park H., Hyun S.W., Kim K.C.;
RT   "Expression of MUC1 mucin gene by hamster tracheal surface epithelial
RT   cells in primary culture.";
RL   Am. J. Respir. Cell Mol. Biol. 15:237-244(1996).
CC   -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN
CC       CYTOSKELETON (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: HIGHLY O-GLYCOSYLATED AND PROBABLY ALSO N-GLYCOSYLATED.
CC   -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U36918; AAB53965.1; -.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF01390; SEA; 1.
DR   SMART; SM00200; SEA; 1.
DR   PROSITE; PS50024; SEA; 1.
KW   Glycoprotein; Signal; Cytoskeleton; Actin-binding; Transmembrane;
KW   Repeat.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    676       MUCIN 1.
FT   DOMAIN       26    582       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    583    603       POTENTIAL.
FT   DOMAIN      604    676       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      458    573       SEA.
FT   CARBOHYD    291    291       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    323    323       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    350    350       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    380    380       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    435    435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    479    479       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    496    496       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    536    536       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   676 AA;  67616 MW;  95F479B6EC5C3884 CRC64;

  Query Match             5.3%;  Score 129;  DB 1;  Length 676;
  Best Local Similarity   20.6%;  Pred. No. 1.6;
  Matches   95;  Conservative   57;  Mismatches  173;  Indels  136;  Gaps   20;

Qy      10 PPV-----TATDSFRTASDASLASSSVRS----VSSDQQREINAIA-DYLTDHVFAAHKL 59
           ||       :||     : | | : ||:| |     | |      : : | :   |  | :
Db     188 PPTSSAVNSATTPVHSGSSAPVTSSAVNSATTPVHSGSSAPVTSSAVNSATTPVHSGSSA 247

Qy      60 PPADSADGQAAVDVHNAQ----ITALIETRASRLHFEGETPATIADTFAKAEKLDRLATT 115
           ||   |     |   ||:       :: :    : :|     ||||              :|
Db     248 PPTSSVVNSATTPVHSGSSAPPTSSAVNLATTPVHSGSSTPAT-------------NST 293

Qy     116 TSGALRATPFAMASLLQYMQ-------PAINKGDWLPAPLKPLTPLISGALSGAMDQVGT 168
           |   |    ||     | :|  :          |: |  :|     | |  | | |||     :
Db     294 TDSA--TTPVPPGSSMQTTEAISGSANTPIHNGSLVPTTSSALVPTTSAAHSGASAMTNS 351

Qy     169 KMMDRAT----------------------GDL-----HYLSASPDRLHD---AMAASV-- 196
           | ||                        | |       | ::   :|:    || :
Db     352 SESDLATTPIDSGTSISTTKAPATTPVHNGSLVPTTSSVLGSATTLIHNDTSTMATTPV 411

Qy     197 ---------KRH--SPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGAVDLGV 245
                     ||  :|:       :: :|: || :     | |:|| :|  | |  | :||
Db     412 GNGTQSSVPSRHPVTPTPPAVSSNSTIALSTYYS-----TALSPAFSSHAAPQ--VSVGV 464

Qy     246 SMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSAS 305
           |              |||    :||   :|      ::| :           |: :|
Db     465 SFF-----------LLSFHIWNHQFNSSL------EDPSSNY---------YQELK--- 494

Qy     306 YSGAALNAGKRMAGLPLDMATDAMGAVRSLV--SASSLTQNGLALAGGFAGVGKL----- 358
                 : ::|| | : : |   : ::   | | :  : :     |     ::
Db     495 ---------RNVSGLFLQVFSRAFLGISTIEFRSGSVVVDSTVIFREGAVNASEVKSQLL 545

Qy     359 ---QEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAAL 396
              ||     |:           |  : | |      ||  | |
Db     546 QHEQEAEEYNLAISKINVGEMQFPSSAQSWPGVPGWGIALL 586


RESULT 6
YAGR_ECO57
ID   YAGR_ECO57     STANDARD;      PRT;   732 AA.
AC   Q8X6J4;
```

```
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative xanthine dehydrogenase yagR, molybdenum binding subunit
DE   (EC 1.1.1.204).
GN   YAGR OR Z0350 OR ECS0314.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.
CC   -!- COFACTOR: Molybdopterin (By similarity).
CC   -!- SUBUNIT: Heterotrimer of yagR, yagS and yagT (Potential).
CC   -!- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE005205; AAG54609.1; -.
DR   EMBL; AP002551; BAB33737.1; -.
KW   Hypothetical protein; Oxidoreductase; Molybdenum; Purine metabolism;
KW   Purine salvage; NAD; Complete proteome.
SQ   SEQUENCE   732 AA;  78135 MW;  A996E91BEB2643F6 CRC64;

  Query Match             5.3%;  Score 129;  DB 1;  Length 732;
  Best Local Similarity   24.7%;  Pred. No. 1.8;
  Matches  104;  Conservative   61;  Mismatches  170;  Indels   86;  Gaps   21;

Qy      22 SDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQITAL 81
           ||| ||: : |:|    :| :  :    :      |:  ||     :   |  : :|||:
Db     250 SDALLAALAARAV----KRPVKVMLPRPSIPNNTTHR--PATLQHLRIGAD-QSGKITAI 302

Qy      82 IETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPFAMASLLQYMQPAINKG 141
              |  : |  |    :|  :  :|       :|| | |   :|:|       : :|
Db     303 -----SHESWSGNLPGGTPETAVQQSEL-----LYAGANRHTGLRLATL------DLPEG 346

Qy     142 DWLPAPLKPLTPLISGALSGAMDQVGTK-----MMDRATGDLHYLSASPDRLHDAMAASV 196
           : : || :   | :  ||  |:|::  |     :  |   |     | | |
Db     347 NAMRAPGE--APGLM-ALEIAIDELAEKAGIDPVEFRILNDTQIDPADPTR--------- 394

Qy     197 KRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGAVDLGVSMAG-GLAANA 255
           | |     : | ||     :  ||              |  | | |  : | |:|  |
Db     395 -RFSRRQLIECLRTGADKFGWKQRN-------------ATPGQVRDGEWLVGHGVA--A 437

Qy     256 GFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGK 315
```

```
           || | ||       :: || |      ||   :: | |  :    | : ||:  |  | :
Db     438 GFRNNLL-------EKSGARV----HLEPNGTVTVETDMTD----IGTGSYTILAQTAAE 482

Qy     316 RMAGLPLDMATDAMGAVRSLVSASSLTQNG--LALAGGFAGVGKLQEMATKNI-TDP-AT 371
           | |:||:     :|     ||| |  | |   : :| :|   ||:||    :  ||  :
Db     483 -MLGVPLEQVAVHLGDSSFPVSAGSGGQWGANTSTSGVYAACVKLREMIASAVGFDPEQS 541

Qy     372 KAAVSQLTNLAGSAAVFAGWTTAALTTDPAVK--------KAESFIQDTVKSTASSTTGY 423
           : |  ::||   || :        || : :::        :  :|    |:    | ||
Db     542 QFADGKITNGTRSAILHEATAGGRLTAEESIEFGTLSKEYQQSTFAGHFVEVGVHSATGE 601

Qy     424 V 424
           |
Db     602 V 602


RESULT 7
SLAP_CAUCR
ID   SLAP_CAUCR     STANDARD;      PRT;  1025 AA.
AC   P35828; Q46015; Q9RFI2;
DT   01-JUN-1994 (Rel. 29, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   S-layer protein (Paracrystalline surface layer protein).
GN   RSAA OR CC1007.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=93007489; PubMed=1393820;
RA   Gilchrist A., Fisher J.A., Smit J.K.;
RT   "Nucleotide sequence analysis of the gene encoding the Caulobacter
RT   crescentus paracrystalline surface layer protein.";
RL   Can. J. Microbiol. 38:193-202(1992).
RN   [2]
RP   REVISIONS TO 376; 636 AND 842-843.
RA   Awram P.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JS3001;
RA   Bingle W.H., Awram P.A., Nomellini J.F., Smit J.K.;
RT   "The secretion signal of C. crescentus S-layer protein is located in
RT   the C-terminal 82 amino acids of the molecule.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RN   [5]
RP   SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=89008089; PubMed=3049545;
RA   Fisher J.A., Smit J.K., Agabian N.;
RT   "Transcriptional analysis of the major surface array gene of
RT   Caulobacter crescentus.";
RL   J. Bacteriol. 170:4706-4713(1988).
RN   [6]
RP   CHARACTERIZATION.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=98292737; PubMed=9620954;
```

```
RA   Awram P., Smit J.K.;
RT   "The Caulobacter crescentus paracrystalline S-layer protein is
RT   secreted by an ABC transporter (type I) secretion apparatus.";
RL   J. Bacteriol. 180:3062-3069(1998).
CC   -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC       OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A
CC       PHYSICAL BARRIER TO PARASITES AND LYTIC ENZYMES.
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER
CC       (TYPE I) SECRETION APPARATUS.
CC   -!- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE
CC       SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A
CC       SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF062345; AAC38665.2; -.
DR   EMBL; AF193063; AAF19365.1; -.
DR   EMBL; AE005779; AAK22991.1; ALT_INIT.
DR   HSSP; P22629; 1SWC.
DR   TIGR; CC1007; -.
DR   InterPro; IPR001343; Hemlysn_Ca_bind.
DR   Pfam; PF00353; hemolysinCabind; 3.
DR   PRINTS; PR00313; CABNDNGRPT.
KW   Cell wall; S-layer; Calcium-binding; Complete proteome.
FT   INIT_MET      0      0
SQ   SEQUENCE   1025 AA;  98001 MW;  AD7A326E1363D8AC CRC64;

  Query Match             5.3%;  Score 128.5;  DB 1;  Length 1025;
  Best Local Similarity   22.4%;  Pred. No. 2.8;
  Matches  114;  Conservative   57;  Mismatches  175;  Indels  163;  Gaps   24;

Qy      12 VTATDSFRTASDASLASSSVRSVSSDQQREINAIADY---------------------- 48
           |:    : ||: | :   |  ::|:|    :|    |
Db     205 VSGIGGYATATAAMINDLSDGALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTA 264

Qy      49 -----LTDHVFAAHKLPPADSADGQAAVDVHN----AQITAL--------IETRASRLHF 91
                :   |  |  |   |:  | |  || |    | :|||        |||   ::
Db     265 NNDTFVAGEVAGAATLTVGDTLSGGAGTDVLNWVQAAAAVTALPTGVTISGIET----MNV 320

Qy      92 EGETPATIADTFAKAEKLDRLATTTSGA-----------LRATPFAMASLLQYMQPAINK 140
               |: :|  :    |  | | ||||             | ||  | |:       |::
Db     321 TSGAAITL-NTSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAA----NNVAVDG 375

Qy     141 GDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHS 200
           |          :|   :|  ||    ||      |:|                :: ||   |
Db     376 G-------ANVTVASTGVTSGT-TTVGAN--SAASG-------------TVSVSVANSS 411

Qy     201 PSLARQVLDTGVAVQT--YSARNAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFG 258
            :    : ||    |   :| |||  | |  |  :          : |:      | |
Db     412 TTTTGAIAVTGGTAVTVAQTAGNAVNTTLTQADVTVTGNSSTTAVTVTQTAAATAGATVA 471

Qy     259 NRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMA 318
           |:          ||: :                    : | : :|  |||
Db     472 GRV----------NGAVTI------------------------TDSAAASATTAGK--- 493

Qy     319 GLPLDMATDAMGAV-RSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQ 377
           :||  :|:   : : :|:||   |:  |   |:|:         :|  ||  | :
Db     494 -----IATVTLGSFGAATIDSSALTTVNLSGTGTSLGIGR------GALT--ATPTANTL 540

Qy     378 LTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQD---TVK---STASSTTGYVADQTVKL 431
             |: |           |||  |:  :|:   |   |:    ||||||   :|
Db     541 TLNVNG-----------LTTTGAITDSEAAADDGFTTINIAGSTASST---IASLVAAD 585

Qy     432 AKTVKDMGGEAIT---HTGASLRN-TVNN 456
           | |:   |   :|   || |:|   || |
Db     586 ATTLNISGDARVTITSHTAAALTGITVTN 614
```

```
RESULT 8
YOPB_YERPS
ID   YOPB_YERPS     STANDARD;      PRT;   401 AA.
AC   Q06114;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   YopB protein.
GN   YOPB.
OS   Yersinia pseudotuberculosis.
OG   Plasmid pIB1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=633;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=YPIII;
RX   MEDLINE=93114907; PubMed=8418066;
RA   Haakansson S., Bergman T., Vanooteghem J.-C., Cornelis G.,
RA   Wolf-Watz H.;
RT   "YopB and YopD constitute a novel class of Yersinia Yop proteins.";
RL   Infect. Immun. 61:71-80(1993).
CC   -!- FUNCTION: PLAYS A ROLE IN VIRULENCE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: TO SHIGELLA INVASION PLASMID IPAB.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L06215; AAA72321.1; -.
KW   Plasmid; Transmembrane; Virulence.
FT   TRANSMEM    166    186       POTENTIAL.
FT   TRANSMEM    187    207       POTENTIAL.
SQ   SEQUENCE   401 AA;  41795 MW;  E434173D7C8B2F64 CRC64;

  Query Match             5.3%;  Score 127.5;  DB 1;  Length 401;
  Best Local Similarity   19.7%;  Pred. No. 1;
  Matches   82;  Conservative   57;  Mismatches  124;  Indels  153;  Gaps   17;

Qy     121 RATPPFAMASLLQYMQ---PA-----------------INKGDWLPAPLKPLTPLISGAL 159
           |:||    ||| |::   ||                    ::|  |||||     :::  :
Db       9 RSTP-VTGSLLPYVETPAPAPLQTQQVAGELKDKNGGVSSQGVQLPAPL----AVVASQV 63

Qy     160 SGAMDQVGTKMMDRATGDL-------HYLS-------ASPD------------------- 186
           :    |  ||::: |           :|:|       ||||
Db      64 TEGQQQEVTKLLESVTRGAAGSQLISNYVSVLTKFTLASPDTFEIELGKLVSNLEEVRKD 123

Qy     187 -------RLHDAMAASVKRHSPSL------ARQVLDTGVAVQTYSARNAVRTVLAPALAS 233
                 |||:     :: :   :      |:||  :|:|  : :   :|: :|:  |:
Db     124 IKIADIQRLHEQNMKKIEENQEKIKETEENAKQVKKSGIASKIFGWLSAIASVIVGAIMV 183

Qy     234 RPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEEND 293
           |  |||       || :   :|                      |:|:
Db     184 ASGV-GAV------AGAMMVASG------------------VIGM-------------- 203

Qy     294 WLEAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFA 353
           |  |:| |:  |       :: |  |      |:  | ::::          ||  |
Db     204 ---ANMAVKQAAEDGLISQEAMKILGPILTAIEVALTVVSTVMTFGGSALKCLANIGAKL 260

Qy     354 GVGKLQEMATKNITDPATKAAVSQ-LTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDT 412
           |      :| |     | | :|  ::|  |||     | : | ||   |::
Db     261 G-ANTASLAAKGAEFSAKVAQISTGISNTVGSAVTKLGGSFAGLTMSHAIR---------- 310

Qy     413 VKSTASSTTGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIE 468
           | | |                    :     |: :  |:|| :|     ||::
Db     311 ---TGSQAT-----------------QVAVGVGSGITQTINNKKQ-----ADLQ 339
```

```
RESULT 9
HIPR_HUMAN
ID   HIPR_HUMAN     STANDARD;      PRT;  1068 AA.
AC   O75146; Q9UED9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Huntingtin interacting protein 1 related (Hip1-related) (Hip 12).
GN   HIP1R OR HIP12 OR KIAA0655.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20515263; PubMed=11063258;
RA   Chopra V.S., Metzler M., Rasper D.M., Engqvist-Goldstein A.E.Y.,
RA   Singaraja R., Gan L., Fichter K.M., McCutcheon K., Drubin D.,
RA   Nicholson D.W., Hayden M.R.;
RT   "HIP12 is a non-proapoptotic member of a gene family including HIP1,
RT   an interacting protein with huntingtin.";
RL   Mamm. Genome 11:1006-1015(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [3]
RP   SEQUENCE OF 179-1068 FROM N.A.
RC   TISSUE=Neuroblastoma;
RX   MEDLINE=99069783; PubMed=9852681;
RA   Seki N., Muramatsu M., Sugano S., Suzuki Y., Nakagawara A., Ohhira M.,
RA   Hayashi A., Hori T., Saito T.;
RT   "Cloning, expression analysis, and chromosomal localization of HIP1R,
RT   an isolog of huntingtin interacting protein (HIP1).";
RL   J. Hum. Genet. 43:268-271(1998).
CC   -!- FUNCTION: COMPONENT OF CLATHRIN-COATED PITS AND VESICLES, THAT MAY
CC       LINK THE ENDOCYTIC MACHINERY TO THE ACTIN CYTOSKELETON.
CC   -!- SUBUNIT: INTERACTS WITH ACTIN AND HUNTINGTIN INTERACTING PROTEIN 1
CC       (HIP1). DOES NOT INTERACT WITH HUNTINGTIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MEMBRANE-ASSOCIATED PROTEIN,
CC       MAINLY LOCALIZED AT THE ENDOCYTIC COMPARTMENTS AND IN THE
CC       PERINUCLEAR REGION (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: BRAIN, HEART, KIDNEY, PANCREAS, AND LIVER, BUT
CC       NOT IN LUNG OR PLACENTA.
CC   -!- DOMAIN: THE TALIN-LIKE DOMAIN BINDS F-ACTIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SLA2 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB014555; BAA31630.1; ALT_INIT.
DR   EMBL; AB013384; BAA33713.1; -.
DR   MIM; 605613; -.
DR   InterPro; IPR001026; ENTH.
DR   InterPro; IPR002558; ILWEQ.
DR   Pfam; PF01417; ENTH; 1.
DR   Pfam; PF01608; I_LWEQ; 1.
DR   ProDom; PD011820; ILWEQ; 1.
DR   SMART; SM00273; ENTH; 1.
DR   SMART; SM00307; ILWEQ; 1.
KW   Actin-binding; Coiled coil.
FT   DOMAIN      347    599       COILED COIL (POTENTIAL).
FT   DOMAIN      800   1068       TALIN-LIKE.
SQ   SEQUENCE   1068 AA;  119388 MW;  3CBC7CF1191BFF8F CRC64;

  Query Match             5.3%;  Score 127;  DB 1;  Length 1068;
  Best Local Similarity   21.4%;  Pred. No. 3.7;
  Matches  117;  Conservative   68;  Mismatches  209;  Indels  154;  Gaps   24;

Qy      33 SVSSDQQREINAIADYLT---DHVFAAHKLPPADSADGQAAVDVHNAQITALIETRASRL 89
           |  :   : |:::  | |:   | :  | :   ||    |: |    |   ||   :
Db     536 SHTEQSKSELSSRLDTLSAEKDALSGAVRQREADLLAAQSLVRETEA---ALSREQQRSS 592

Qy      90 HFEGETPATIADTFAKAEKL-DRLATTTSGALRATPFAMASLLQYMQPAINKGDWLPAPL 148
             :||    :|:  :: : |  ||        ||       | :|   | |::| |
Db     593 QEQGELQGRLAERESQEQGLRQRLLDEQFAVLRGAAAEAAGIL---QDAVSKLD------ 643

Qy     149 KPL------TP--LISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHS 200
           ||       :|  |:| |   |:| | |        |    ||::  |    |: |:: | |
Db     644 DPLHLRCTSSPDYLVSRA-QEALDAVST----LEEGHAQYLTSLADA--SALVAALTRFS 696

Qy     201 PSLARQVLDTGVAVQTYSARNAVRTV---------------------LAPALAS--RPA 236
             |  ::: |          | | :                      |    ||  |
Db     697 HLAADTIINGGATSHLAPTDPADRLIDTCRECGARALELMGQLQDQQALRHMQASLVRTP 756

Qy     237 VQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLE 296
           :|| : ||  :          :  |:  |  : |  :     |||  |  :   | :
Db     757 LQGILQLGQEL-----------KPKSLDVRQEELGAVV-----DKEMAATSAAIEDAVR 799

Qy     297 AYKAIKSA---SYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFA 353
           : : :    : ||  |   :|:    |: || | |:| ||: |:  |  :  :|
Db     800 RIEDMMNQARHASSGVKLEVNERI----LNSCTDLMKAIRLLVTTSTSLQKEIVESG--R 853

Qy     354 GVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTV 413
           |    ||   ||          |: |    ||:   ||    |   .   |: |:  :  |
Db     854 GAATQQEFYAKN----------SRWTEGLISASKAVGWGATQL-----VEAADKVVLHTG 898

Qy     414 K---------STASSTTGYVADQTVK------------LAKTVKDMGGEAITHT----- 446
           |           |:||   ||   ||              ::|| :     :  |
Db     899 KYEELIVCSHEIAASTAQLVAASKVKANKHSPHLSRLQECSRTVNERAANVVASTKSGQE 958

Qy     447 ----------GASLRNTVNNLRQRPAREADIEE------------------GGTAASP 476
                    | ||       :   |  ::|:                    | : ||
Db     959 QIEDRDTMDFSGLSLIKLKKQEMETQVRVLELEKTLEAERMRLGELRKQHYVLAGASGSP 1018

Qy     477 -SEIPFRP 483
            |:  ||
Db    1019 GEEVAIRP 1026


RESULT 10
SBCC_RHOCA
ID   SBCC_RHOCA     STANDARD;      PRT;  1238 AA.
AC   O68032;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nuclease sbcCD subunit C.
GN   SBCC.
OS   Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1061;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SB1003 / St Louis;
RX   MEDLINE=97404404; PubMed=9256491;
RA   Vlcek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;
RT   "Sequence of a 189-kb segment of the chromosome of Rhodobacter
RT   capsulatus SB1003.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).
CC   -!- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
CC       can inhibit DNA replication and are intermediates in certain DNA
```

```
CC       recombination reactions. The complex acts as a 3'->5' double
CC       strand exonuclease that can open hairpins. It also has a 5'
CC       single-strand endonuclease activity (By similarity).
CC    -!- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).
CC    -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; AF010496; AAC16118.1; -.
DR    InterPro; IPR003439; ABC_transportr.
KW    Hydrolase; Nuclease; Exonuclease; Endonuclease; DNA replication;
KW    DNA recombination; ATP-binding; Coiled coil.
FT    NP_BIND      37     44       ATP (POTENTIAL).
FT    DOMAIN      395    438       COILED COIL (POTENTIAL).
FT    DOMAIN      466    487       COILED COIL (POTENTIAL).
FT    DOMAIN      521    600       COILED COIL (POTENTIAL).
FT    DOMAIN      724    770       COILED COIL (POTENTIAL).
FT    DOMAIN      901    943       COILED COIL (POTENTIAL).
FT    DOMAIN     1019   1052       COILED COIL (POTENTIAL).
SQ    SEQUENCE   1238 AA;  128046 MW;  2B6BC6C63A859AE1 CRC64;

  Query Match             5.3%;  Score 127;  DB 1;  Length 1238;
  Best Local Similarity   22.6%;  Pred. No. 4.4;
  Matches  131;  Conservative   76;  Mismatches  214;  Indels  158;  Gaps   28;

Qy      14 ATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDV 73
           | ::   |:|::||    :: :: ||    |:|     : |||  ||| :||  | ::
Db     486 AAEARDQAADSALA----KAAAALQQ----ALAP-----LEAAH--PPARTAD-LARIET 529

Qy      74 HNAQITALIE--TRASRLHFEGETPATIADT---FAKAEK------LDRLATTTSG---- 118
             |::|  :     |: |    :  || | |    |:|::      |||  |  :
Db     530 DLAELTRALRDGAEAAALGAAAQRAATAAATEAALARAQESAAKQDLDRAETQIAALTAP 589

Qy     119 ------------ALRA-----TPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALS 160
                       :|||     :|  :   |:: ||      |   |:       ||
Db     590 LEQADLALSDAARSLRAQLSAGSPCPVCGALEHPTPAEAGLAHLAERLRADQAAARGAAQ 649

Qy     161 GAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSAR 220
           | | :      ||| :     |: |:      |   | :          |       |||
Db     650 AARDALTAAQGARATAEARGTQAAEDQRRAQTRAEAARAA---------WGDTQPRVSAR 700

Qy     221 NAVRTVLAPALASRP---AVQGAVD----LGVSMAGGLAANAGFGNRLLSVQSRDHQRGG 273
                 |||||    |   |:  | |     |  : |   |  :    ||   : || :|
Db     701 -----PLAPALPGTPDPTALAAAQDRLCALQTAEAAAQAEISALRTRLTEAE-RDRERLR 754

Qy     274 ALVL---GLKDK--EPKAQLSEENDWLEAYKAIKSASYSGAAL-------NAGKRMAGLP 321
             :|    | :::    :|: ::|     || :   :|   | ||         ||:     |
Db     755 RALLAHRGTRERLAVQQAETAQEAALAEARRTEAAARRDGLALALAPALARAGEDDPAAP 814

Qy     322 -----LDMATDAMGAVRSLVSASSLTQNGL-ALAGGFAGVGKLQEMATKNITDPATKA-- 373
                |      |:|| |: : |:    |  | |||    |   :  | ||       |  |
Db     815 GLAERLAATVSAVGAARTGLQAA---QEALSALAPQLAAARRDSETATAQAQSAAQAARD 871

Qy     374 ---AVSQLTNLAGSAAVFAGWTTA----------------------ALTTDPAVKKAES 407
             | : |    |  | :  |  ||                       |   |:  ||:
Db     872 RDGAWAALR--AERAPLLDGQPTALHRSRFNDQRLAAQRAQQAAAADLATAQAALAAAEA 929

Qy     408 FIQDTVKSTASSTTGY-------------VADQTVKLAKTV-----------KDMGG--E 441
             :| :: : : |                 |    :||  :           |: |  :
Db     930 RAAETARAASEAATAQRAAEADLAAALEAAAMSAAELAALIALPPGTAERLRADLRGRDD 989

Qy     442 AITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSEIP 480
           |:|   ::|           ||: |::|     :|:| |
Db     990 AVTAARSAL----------GARKTDLQEIEDQGNPAEDP 1018


RESULT 11
```

```
NCCB_ALCXX
ID   NCCB_ALCXX     STANDARD;      PRT;   397 AA.
AC   Q44585;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Nickel-cobalt-cadmium resistance protein nccB.
GN   NCCB.
OS   Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxidans).
OG   Plasmid pTOM9.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Achromobacter.
OX   NCBI_TaxID=515;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=31A;
RX   MEDLINE=95050278; PubMed=7961470;
RA   Schmidt T., Schlegel H.G.;
RT   "Combined nickel-cobalt-cadmium resistance encoded by the ncc locus
RT   of Alcaligenes xylosoxidans 31A.";
RL   J. Bacteriol. 176:7045-7054(1994).
CC   -!- FUNCTION: COMPONENT OF THE NCC CATION-EFFLUX SYSTEM THAT CONFERS
CC       RESISTANCE TO NICKEL, COBALT AND CADMIUM.
CC   -!- SIMILARITY: HIGH, TO A.EUTROPHUS CNRB AND SOME, TO ALCALIGENES
CC       CZCB.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L31363; AAA65105.1; -.
DR   InterPro; IPR002215; HlyD.
DR   Pfam; PF00529; HlyD; 1.
DR   TIGRFAMs; TIGR00999; 8a0102; 1.
KW   Plasmid; Nickel; Cobalt; Cadmium; Transport.
SQ   SEQUENCE   397 AA;  39933 MW;  C6DE57CE314996B8 CRC64;

  Query Match             5.2%;  Score 126.5;  DB 1;  Length 397;
  Best Local Similarity   23.2%;  Pred. No. 1.2;
  Matches   88;  Conservative   59;  Mismatches  153;  Indels   79;  Gaps   18;

Qy     149 KPLTPLISGALS-------GAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSP 201
           :|  |:|:|  :         ||   :|:     :  ::  |:|:|::      |||     |
Db       9 QPSWPMIAGVAAAAALVGFGAARGLGSP----SGAEVSKLAAAPEK----AAASAPAAEP 60

Qy     202 SLAR----QVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGF 257
           :  |     :    :||:  ||      :|||  ||  || |:  :  | | |
Db      61 AEVRIPGEYLAAANIAVEPVSAGGVGSVLLAP--ASVAAVPGSEAVIASRAAG------- 111

Qy     258 GNRLLSVQSR--DHQRGGALVLGLKDKEPKAQLSEENDWLEA--------YKAIKSASYS 307
              :| :| :  |  | |  || | |     | :: |     :|         |:   |
Db     112 --AVLRIQRKLGDAVRAGD-VLALVDSPEAAAMAAERKVAQARADLARKTYERESSLFQQ 168

Qy     308 GAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNG--LALAGGFAGVGKLQEMATKN 365
           |          | : ||:|    :     ::  |: :: :|  :|:     ||     :  ::
Db     169 GVTPRQEMESARIALDVAQAEVQRAATVAQAAKVSSDGRSVAVVSPIAG-----RITAQS 223

Qy     366 ITDPATKAAVSQLTNLAGSAAV-FAGWTTAALT------TDPAVKKAESF-IQDTVKSTA 417
           :|  |  |  ::|  :|||  ||     : ||| |       :|  :  |     :   |::
Db     224 VTLGAYVAPQAELFRVAGSGAVQVEAYVTAADTSRIAAGSDATIVLANGAPLAGRVQAVT 283

Qy     418 SSTTGYVADQTVKLAKTVKDMG---GEAI---THTGASLRN-------TVNNLRQRPARE 464
             : :|     || :     : |   || :    || |:  |       | ||   |
Db     284 PTVSGSARAATVVVTPVDANSGLIVGEGVQVRLHTKAADANAMSVPEDAVQNLDGRDVVF 343

Qy     465 ADIEEGGTAASPSEIPFRP 483
             ::|          |||
Db     344 VRTQQG----------FRP 352
```

```
RESULT 12
FLIC_SALRU
ID   FLIC_SALRU     STANDARD;      PRT;   492 AA.
AC   P06175;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Flagellin (Phase-1-R flagellin).
GN   FLIC OR FLAF OR HAG.
OS   Salmonella rubislaw.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=598;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 10717;
RX   MEDLINE=87040788; PubMed=3774559;
RA   Wei L.-N., Joys T.M.;
RT   "The nucleotide sequence of the H-1r gene of Salmonella rubislaw.";
RL   Nucleic Acids Res. 14:8227-8227(1986).
CC   -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC       FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC   -!- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC       BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC       PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04505; CAA28190.1; -.
DR   PIR; S07375; S07375.
DR   InterPro; IPR001492; FlagellinN.
DR   InterPro; IPR001029; Flagellin_C.
DR   Pfam; PF00669; Flagellin_N; 1.
DR   Pfam; PF00700; Flagellin_C; 1.
DR   PRINTS; PR00207; FLAGELLIN.
DR   ProDom; PD000316; Flagellin_C; 1.
KW   Flagella.
FT   INIT_MET      0      0
SQ   SEQUENCE   492 AA;  51285 MW;  0F87CB00D311B5FA CRC64;

  Query Match             5.2%;  Score 125.5;  DB 1;  Length 492;
  Best Local Similarity   21.9%;  Pred. No. 1.7;
  Matches  110;  Conservative   56;  Mismatches  192;  Indels  145;  Gaps   20;

Qy      63 DSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL-- 120
           | | |||  :   | |  |  |:|||   :| :                :| || |||
Db      42 DDAAGQAIANRFTANIKGL--TQASRNANDGIS---------------IAQTTEGALNE 83

Qy     121 ------RATPFAMASL--------LQYMQPAINK--------------GDWLPAPLKPL 151
                |    |: |          |  :|  | :              |  : |    |
Db      84 INNNLQRVRELAVQSANSTNSQSDLDSIQAEITQRLNEIDRVSGQTQFNGVKVLAQDNTL 143

Qy     152 TPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLARQVLDTG 211
           |  : ||  |    :  | ::  |  |  |:          ||:|  :: |       |
Db     144 TIQV-GANDGETIDIDLKQINSQTLGLDTLNVQQKYKVSDTAATVTGYTDSA------TA 196

Qy     212 VAVQTYSARNAVRTVLAPALASRPAVQGAVDLG------VSMAGGLAANAGFGNRLLSVQ 265
           :  |::|         |  |   ||: | :          :  |  |  |     ::
Db     197 IDKSTFAAS-------ATTLGGTPAITGDLKFDDTTGKYYADVSGTTAKDGVYEVTVAAD 249

Qy     266 SRDHQRG---GALVLGLKD---KEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMAG 319
             :   |   | :  |       |: |     |  |  ||  |: :|  : |
Db     250 GKVTLTGTPTGPITAGFPSTATKDVKQTQQENADLTEAKAALTAAGVAAA---------- 299

Qy     320 LPLDMATDAMGAVRSLVSASSLTQNGLALAGGFA------------GVGKLQEMATKNI 366
```

```
                    ||:|  |    ||  : || |                | :    ||
Db     300 ----------GHRSVVKMSYTDNNGKTIDGGLAVKVGDDYYSATQNKDGSISINTTKYT 348

Qy     367 TDPAT-KAAVSQLTNLAGSAAV--FAGWTTAALTTDPAVKKAESFIQDTVKSTASSTTGY 423
           |  | | |:::|    |   |    | | ||   :    ||:    |  :: |::|
Db     349 ADNGTSKTALNKLGGADGKTEVVSIGGKTYAASKAEGHNFKAQ---PDLAEAAATTTENP 405

Qy     424 VADQTVKLAK---------TVKDMGGEAITHTGASLRNTVNNLRQRPAREAD-------- 466
           :     ||:           |::    |||: |    ||||||    :|  |
Db     406 LQKIDAALAQVDTLRSDLGAVQNRFNSAITNLG----NTVNNLTSARSRIEDSDYATEVS 461

Qy     467 -------IEEGGTA--ASPSEIP 480
                  ::: ||:  |  :::|
Db     462 NMSRAQILQQAGTSVLAQANQVP 484


RESULT 13
SED4_YEAST
ID   SED4_YEAST     STANDARD;      PRT;  1065 AA.
AC   P25365;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   SED4 protein.
GN   SED4 OR YCR067C OR YCR67C OR YCR901.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE OF 1-514 FROM N.A.
RA   Contreras R., Demolder J., Fiers W., Molemans F.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 446-1065 FROM N.A.
RX   MEDLINE=92221691; PubMed=1561837;
RA   Benit P., Chanet R., Fabre F., Faye G., Fukuhara H., Sor F.;
RT   "Sequence of the sup61-RAD18 region on chromosome III of
RT   Saccharomyces cerevisiae.";
RL   Yeast 8:147-153(1992).
RN   [3]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=93011014; PubMed=1327759;
RA   Hardwick K.G., Boothroyd J.C., Rudner A.D., Pelham H.R.B.;
RT   "Genes that allow yeast cells to grow in the absence of the HDEL
RT   receptor.";
RL   EMBO J. 11:4187-4195(1992).
CC   -!- FUNCTION: WHEN PRESENT IN MULTIPLE COPIES, ALLOWS CELLS TO GROW
CC       IN THE ABSENCE OF HDEL RECEPTOR BY EITHER SLOWING TRANSPORT FROM
CC       THE ER OR BY STIMULATING VESICLE BUDDING FROM GOLGI MEMBRANES.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein.
CC   -!- MISCELLANEOUS: IN THE PROCESS OF TRANSPORT, SED4 ITSELF MAY
CC       MIGRATE TO THE GOLGI APPARATUS AND FUNCTION IN SUBSEQUENT
CC       TRANSPORT EVENTS.
CC   -!- SIMILARITY: CONTAINS 2 WD REPEATS (TRP-ASP DOMAINS).
CC   -!- SIMILARITY: IN THE N-TERMINAL HALF, TO YEAST SEC12.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59720; CAA42273.1; -.
DR   PIR; S19482; S19482.
DR   SGD; S0000663; SED4.
DR   InterPro; IPR000886; ER_target.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 2.
DR   SMART; SM00320; WD40; 2.
DR   PROSITE; PS00014; ER_TARGET; 1.
```

```
DR   PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
DR   PROSITE; PS50082; WD_REPEATS_2; FALSE_NEG.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Transport; Protein transport; Glycoprotein; Endoplasmic reticulum;
KW   Transmembrane; Signal-anchor; Golgi stack; Repeat; WD repeat.
FT   DOMAIN        1    346       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    347    365       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN      366   1065       LUMENAL (POTENTIAL).
FT   REPEAT      259    298       WD 1.
FT   REPEAT      302    341       WD 2.
FT   DOMAIN      467    476       POLY-SER.
FT   DOMAIN      579    590       POLY-SER.
FT   DOMAIN      824    863       4 X 10 AA TANDEM REPEATS.
FT   REPEAT      824    833       1.
FT   REPEAT      834    843       2.
FT   REPEAT      844    853       3.
FT   REPEAT      854    863       4.
FT   CARBOHYD    388    388       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1039   1039       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   SITE        1062   1065      PREVENT SECRETION FROM ER (POTENTIAL).
SQ   SEQUENCE   1065 AA;  114079 MW;  0F1DAE0D428085A8 CRC64;

  Query Match             5.2%;  Score 125.5;  DB 1;  Length 1065;
  Best Local Similarity  19.8%;  Pred. No. 4.5;
  Matches   90;  Conservative   93;  Mismatches  171;  Indels  101;  Gaps   22;

Qy      12 VTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAV 71
           :: : |  |:   :|::|:  :|:: |    |  |:  |  :         |::   :
Db     585 LSLSSSLTTSPTTALSTSTATAVTTTQTNPTNDAAN--TSFL---------DNSKPASTR 638

Qy      72 DVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGALRATPF-AMASL 130
           :::  :|   : |:    :      ||  :|: |:||:      ||| ::  ||   | |
Db     634 EIYKTKIITEVITKIEYRNI----PA--SDSNAEAEQY----VTTSSSMLLTPTDTMVS- 682

Qy     131 LQYMQPAINKGDWLPAPLKPL--TPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRL 188
                 ::: | : : |: :  ||  | :::   | | : ::     ::   ||| |
Db     683 -----SPVSEIDPIASELERMVETPTHSISIASEFDSVASNLIPNE--EILSTSASQD-- 733

Qy     189 HDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASR--PAVQG------- 239
                 |:  |  : :   : :|   |:      | || :  |||   |::
Db     734 ------SISSHPSTFSDSSITSG--FQSIE----VSTVTSSVLASESIPSISDSTFSKFH 781

Qy     240 AVDLGVSMAGGLAANAGFG------NRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEEND 293
           ::   || |    | : |        :|:::  : | :|  ||:    |      :  :|
Db     782 SISEPVSSAIVETATSSFSKTETKTSRVIAFSTEDSERSSALI----DNSEYTSVLADN- 836

Qy     294 WLEAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQN--------- 344
            ||    :   |   : |          |  : |||: : :: |  |||:::
Db     837 -LEPTSVLADNSEPTSVLADSSE----PTSVFTDAVQSPKTSVGQSSLSESTNIEGTSMA 891

Qy     345 GLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKK 404
            :  :   | :| | ::    ::  :  : |:|           | ||
Db     892 SMIFSSSGASIGALSDIGKGTLSVESASSTVAQ---------PMPGVTTT---------- 932

Qy     405 AESFIQDTVKSTASS--TTGYVADQTVKLAKTVKD 437
           | ||:    | :|||   :|:|  : :   :: ||
Db     933 APSFVSSPHKISASSIDASGFVQKEIMIEVQSSKD 967


RESULT 14
HIPR_MOUSE
ID   HIPR_MOUSE     STANDARD;      PRT;  1068 AA.
AC   Q9JKY5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Huntingtin interacting protein 1 related (Hip1-related).
GN   HIP1R.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20082888; PubMed=10613908;
RA   Engqvist-Goldstein A.E.Y., Kessels M.M., Chopra V.S., Hayden M.R.,
RA   Drubin D.G.;
RT   "An actin-binding protein of the Sla2/Huntingtin interacting protein 1
RT   family is a novel component of clathrin-coated pits and vesicles.";
RL   J. Cell Biol. 147:1503-1518(1999).
CC   -!- FUNCTION: COMPONENT OF CLATHRIN-COATED PITS AND VESICLES, THAT MAY
CC       LINK THE ENDOCYTIC MACHINERY TO THE ACTIN CYTOSKELETON.
CC   -!- SUBUNIT: INTERACTS WITH ACTIN AND HUNTINGTIN INTERACTING PROTEIN 1
CC       (HIP1). DOES NOT INTERACT WITH HUNTINGTIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MEMBRANE-ASSOCIATED PROTEIN,
CC       MAINLY LOCALIZED AT THE ENDOCYTIC COMPARTMENTS AND IN THE
CC       PERINUCLEAR REGION.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. EXPRESSED AT LOWER LEVELS IN
CC       SKELETAL MUSCLE AND HEART. THE LEVEL OF EXPRESSION DOES NOT CHANGE
CC       APPRECIABLY DURING DEVELOPMENT.
CC   -!- DOMAIN: THE TALIN-LIKE DOMAIN BINDS F-ACTIN.
CC   -!- SIMILARITY: BELONGS TO THE SLA2 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF221713; AAF34662.1; -.
DR   MGD; MGI:1352504; Hip1r.
DR   InterPro; IPR001026; ENTH.
DR   InterPro; IPR002558; ILWEQ.
DR   Pfam; PF01417; ENTH; 1.
DR   Pfam; PF01608; I_LWEQ; 1.
DR   ProDom; PD011820; ILWEQ; 1.
DR   SMART; SM00273; ENTH; 1.
DR   SMART; SM00307; ILWEQ; 1.
KW   Actin-binding; Coiled coil.
FT   DOMAIN        77     80       POLY-SER.
FT   DOMAIN       311    316       POLY-GLU.
FT   DOMAIN       346    644       COILED COIL (POTENTIAL).
FT   DOMAIN       800   1068       TALIN-LIKE.
SQ   SEQUENCE   1068 AA;  119484 MW;  97CEE9D92CDF5DB1 CRC64;

  Query Match             5.2%;  Score 125.5;  DB 1;  Length 1068;
  Best Local Similarity  20.9%;  Pred. No. 4.5;
  Matches  117;  Conservative   72;  Mismatches  206;  Indels  165;  Gaps   24;

Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQI 78
           ||    |  || : :::::::    |:: :           |:  ||  :   :
Db     537 RTEQSGSELSSRLDTLNAEKE----ALSGVVRQR----------EAELLAAQSLVREKE 581

Qy      79 TALIETRASRLHFEGETPATIADTFAKAEKL------DRLATTTSGALRATPFAMASLLQ 132
           || : :      :||    :|:  :: : |      ::||   | |  |
Db     582 EALSQEQQRSSQEKGELRGQLAEKESQEQGLRQKLLDEQLAVLRSAAAEAEAI------- 634

Qy     133 YMQPAINKGDWLPAPLKPL------TP--LISGALSGAMDQVGTKMMDRATGDLHYLSAS 184
            :| |::| |      ||      :|  |:| |   |:| |         |   ||::|
Db     635 -LQDAVSKLD------DPLHLRCTSSPDYLVSRA-QAALDSVS----GLEQGHTQYLASS 682

Qy     185 PDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAV---------------------- 223
           |     |: |:: | |  ||   :  | |     :  :
Db     683 EDA--SALVAALTRFS-HLAADTIVNGAATSHLAPTDPADRLMDTCRECGARALELVGQL 739

Qy     224 --RTVLAPALAS--RPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGL 279
             :|||  |  |  |  :|| : ||  :            :  |:  |  : |      :
Db     740 QDQTVLRRAQPSLMRAPLQGILQLGQDL-----------KPKSLDVRQEELG-----AM 782

Qy     280 KDKEPKAQLSEENDWLEAYKAIKSAS---YSGAALNAGKRMAGLPLDMATDAMGAVRSLV 336
           |||  |  :   | :   : : | :    ||  |   :|:    |:  || | |:| ||
Db     783 VDKEMAATSAAIEDAVRRIEDMMSQARHESSGVKLEVNERI----LNSCTDLMKAIRLLV 838
```

```
Qy     337 SASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAAL 396
           |:  |  :  :|   |     ||   ||          |: |    ||:   ||     |
Db     839 MTSTSLQKEIVESG--RGAATQQEFYAKN----------SRWTEGLISASKAVGWGATQL 886

Qy     397 --TTDPAVKKAESFIQDTVKS--TASSTTGYVADQTVK-----------LAKTVKDMG 439
             : |  |     : :  | |   |:||   ||   ||              ::|| :
Db     887 VESADKVVLHMGKYEELIVCSHEIAASTAQLVAASKVKANKNSPHLSRLQECSRTVNERA 946

Qy     440 GEAITHT---------------GASLRNTVNNLRQRPAREADIEE------------- 469
           :  |                  | ||        :   |  ::|:
Db     947 ANVVASTKSGQEQIEDRDTMDFSGLSLIKLKKQEMETQVRVLELEKTLEAERVRLGELRK 1006

Qy     470 ------GGTAASPSEIPFRP 483
                ||      | | ||
Db    1007 QHYVLAGGMGTPSEEEPSRP 1026


RESULT 15
TALI_MOUSE
ID   TALI_MOUSE     STANDARD;      PRT;  2541 AA.
AC   P26039;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Talin.
GN   TLN1 OR TLN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Fibroblast;
RX   MEDLINE=91015390; PubMed=2120593;
RA   Rees D.J.G., Ades S.A., Singer S.J., Hynes R.O.;
RT   "Sequence and domain structure of talin.";
RL   Nature 347:685-689(1990).
CC   -!- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL
CC       STRUCTURES TO THE PLASMA MEMBRANE. HIGH MOLECULAR WEIGHT
CC       CYTOSKELETAL PROTEIN CONCENTRATED AT REGIONS OF CELL-SUBSTRATUM
CC       CONTACT AND, IN LYMPHOCYTES, AT CELL-CELL CONTACTS.
CC   -!- SUBUNIT: BINDS WITH HIGH AFFINITY TO VINCULIN AND WITH LOW
CC       AFFINITY TO INTEGRINS.
CC   -!- PTM: PHOSPHORYLATED.
CC   -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC   -!- SIMILARITY: IN THE C-TERMINUS TO YEAST SLA2 AND C.ELEGANS ZK370.3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56123; CAA39588.1; -.
DR   PIR; S11661; S11661.
DR   MGD; MGI:1099832; Tln.
DR   InterPro; IPR000299; Band_4.1.
DR   InterPro; IPR002558; ILWEQ.
DR   InterPro; IPR000159; RA_domain.
DR   Pfam; PF00373; Band_41; 1.
DR   Pfam; PF01608; I_LWEQ; 1.
DR   ProDom; PD011820; ILWEQ; 1.
DR   SMART; SM00295; B41; 1.
DR   SMART; SM00307; ILWEQ; 1.
DR   PROSITE; PS00660; BAND_41_1; 1.
DR   PROSITE; PS00661; BAND_41_2; 1.
DR   PROSITE; PS50057; BAND_41_3; 1.
KW   Structural protein; Cytoskeleton; Phosphorylation.
FT   DOMAIN      173    332       BAND 4.1-LIKE.
FT   VARIANT    1105   1105       L -> P.
FT   VARIANT    2180   2180       K -> M.
SQ   SEQUENCE   2541 AA;  269832 MW;  14EF75ABE9FC2CBB CRC64;

  Query Match             5.2%;  Score 125.5;  DB 1;  Length 2541;
  Best Local Similarity   22.2%;  Pred. No. 13;
  Matches  129;  Conservative   69;  Mismatches  242;  Indels  141;  Gaps   26;

Qy       8 QQPP---VTATDSFRTASDA-SLASSSVRSVSSDQ---QREINAIADYLTD-----HVFA 55
           : ||    | |  | || ||:     :|:|  :  |    :  | |:   | |
Db    1625 RDPPRWSVLAGHS-RTVSDSIKKLITSMRDKAPGQLECETAIAALNSCLRDLDQASLAAV 1683

Qy      56 AHKLPPADSADGQAAVDVHNAQITA------LIETRASRLHFEGE-------------TP 96
           : :| | :    :|   :|   :||      |||  ||    |                 |
Db    1684 SQQLAPREGISQEA---LHTQMLTAVQEISHLIEPLASAARAEASQLGHKVSQMAQYFEP 1740

Qy      97 ATIADTFAKAEK---------LDRLATTTSGALRA---------------TPFAMASLL 131
           . |:|   | ::          ||:  |    ||:                   |  |:   :
Db    1741 LTLAAVGAASKTLSHPQQMALLDQTKTLAESALQLLYTAKEAGGNPKQAAHTQEALEEAV 1800

Qy     132 QYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDA 191
           | |  |:          : ||  ::  | | |   |   |:|  |  :: |   |  : |
Db    1801 QMMTEAV----------EDLTTTLNEAASAA--GVVGGMVDSITQAINQLDEGP--MGDP 1846

Qy     192 MAASVKRHSPSLARQVLDTGVAVQTYSAR-NAVRTVLAPALASRPAVQGAVDLGVSMAGG 250
             : |  :  :: |      | ||   : |    | |     :  | :      |
Db    1847 EGSFVD-YQTTMVRTAKAIAVTVQEMVTKSNTSPEELGPLANQLTSDYGRLASQAKPAAV 1905

Qy     251 LAANAGFGNRLL-SVQSRDH------QRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKS 303
           | |   |  :   ||   |       : |||     |   | :| |      | :  :
Db    1906 AAENEEIGAHIKHRVQELGHGCSALVTKAGALQCSPSDVYTKKELIE-----CARRVSEK 1960

Qy     304 ASYSGAALNAGKR----------MAGLPLDMATDAMGAVRSLVSASSLTQNGLALAG-- 350
           |:  ||| || |            ::|:  |: |  | |     :| :| : |
Db    1961 VSHVLAALQAGNRGTQACITAASAVSGIIADLDTTIMFA-----TAGTLNREGAETFADH 2015

Qy     351 --GFAGVGKLQEMATKNITD----------PATKAAVSQLTNLA-----GSAAVFAGWTT 393
              |     |:    || :            | :::|: :| ||     |:|:: |
Db    2016 REGILKTAKVLVEDTKVLVQNAAGSQEKLAQAAQSSVATITRLADVVKLGAASLGAEDPE 2075

Qy     394 AALTTDPAVKKAESFIQDTVKSTASSTTGYVAD---------------QTVKLAKTVKD 437
             :    |||     : | : :|  :  | | |                   | ||||
Db    2076 TQVVLINAVKDVAKALGDLISAT-KAAAGKVGDDPAVWQLKNSAKVMVTNVTSLLKTVKA 2134

Qy     438 MGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAASPSE 478
           :  ||   | | |  |  ::||  |     |     ::| :
Db    2135 VEDEATKGTRA-LEATTEHIRQELAVFCSPEPPAKTSTPED 2174


Search completed: January  2, 2003, 15:19:19
Job time : 16.4902 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:17:36 ; Search time 13.4861 Seconds
(without alignments)
1060.315 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHINRRVQQPPVTATDSFRT.....IEEGGTAASPSEIPFRPMRS 486

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	130.5	5.4	1996	2	US-08-804-227C-9	Sequence 9, Appli
2	130.5	5.4	1996	2	US-08-804-198-3	Sequence 3, Appli
3	128.5	5.3	1026	1	US-08-194-290-7	Sequence 7, Appli
4	127.5	5.3	401	2	US-08-591-079-6	Sequence 6, Appli
5	127	5.3	676	4	US-09-085-199B-7	Sequence 7, Appli
6	127	5.3	2123	4	US-08-968-685A-10	Sequence 10, Appl
7	126.5	5.2	669	4	US-09-071-035-264	Sequence 264, App
8	126.5	5.2	1026	2	US-08-614-377A-7	Sequence 7, Appli
9	126.5	5.2	1026	4	US-09-142-648B-7	Sequence 7, Appli
10	126.5	5.2	1638	4	US-09-071-035-258	Sequence 258, App
11	126.5	5.2	1638	4	US-09-071-035-262	Sequence 262, App
12	126.5	5.2	1638	4	US-09-071-035-266	Sequence 266, App
13	125.5	5.2	1068	4	US-09-085-199B-11	Sequence 11, Appl
14	123	5.1	1024	4	US-09-562-737-41	Sequence 41, Appl
15	122	5.1	488	4	US-08-809-326A-1	Sequence 1, Appli
16	122	5.1	649	4	US-08-809-326A-15	Sequence 15, Appl
17	122	5.1	5215	4	US-09-105-537-2	Sequence 2, Appli
18	118.5	4.9	391	4	US-08-818-112-107	Sequence 107, App
19	118.5	4.9	391	4	US-08-818-111-102	Sequence 102, App
20	118.5	4.9	391	4	US-09-056-556-107	Sequence 107, App
21	118.5	4.9	391	4	US-09-072-596-102	Sequence 102, App
22	118.5	4.9	1024	4	US-09-562-737-44	Sequence 44, Appl
23	118	4.9	2285	4	US-09-308-375-2	Sequence 2, Appli
24	117.5	4.9	1024	4	US-09-562-737-43	Sequence 43, Appl
25	116	4.8	1024	4	US-09-562-737-45	Sequence 45, Appl
26	115	4.8	1832	3	US-09-335-409-4	Sequence 4, Appli
27	115	4.8	1832	4	US-09-568-102-4	Sequence 4, Appli
28	115	4.8	1832	4	US-09-567-969-4	Sequence 4, Appli
29	115	4.8	1832	4	US-09-568-480-4	Sequence 4, Appli
30	115	4.8	1832	4	US-09-568-486-4	Sequence 4, Appli
31	115	4.8	1832	4	US-09-568-472-4	Sequence 4, Appli
32	115	4.8	1832	4	US-09-567-899-4	Sequence 4, Appli
33	115	4.8	3031	1	US-07-689-008-2	Sequence 2, Appli
34	114.5	4.7	748	4	US-09-066-046-4	Sequence 4, Appli
35	114.5	4.7	2756	1	US-08-375-709-11	Sequence 11, Appl
36	114.5	4.7	2756	1	US-08-752-929-11	Sequence 11, Appl
37	114.5	4.7	2756	4	US-09-090-793-7	Sequence 7, Appli
38	114	4.7	688	3	US-09-141-047-8	Sequence 8, Appli
39	114	4.7	7257	3	US-09-335-409-5	Sequence 5, Appli
40	114	4.7	7257	4	US-09-568-102-5	Sequence 5, Appli
41	114	4.7	7257	4	US-09-567-969-5	Sequence 5, Appli
42	114	4.7	7257	4	US-09-568-480-5	Sequence 5, Appli
43	114	4.7	7257	4	US-09-568-486-5	Sequence 5, Appli
44	114	4.7	7257	4	US-09-568-472-5	Sequence 5, Appli
45	114	4.7	7257	4	US-09-567-899-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-804-227C-9
; Sequence 9, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1996 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-804-227C-9

Query Match 5.4%; Score 130.5; DB 2; Length 1996;
Best Local Similarity 21.5%; Pred. No. 0.008;
Matches 133; Conservative 70; Mismatches 216; Indels 199; Gaps 28;

```
Qy      12 VTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHV--------------FAAH 57
           :| | : |||||     | :  | :  :| : |:   | |                 ||
Db    1342 LTTTAAGRTASDGG-RLSLLDEVRATVRRTLEAVQARLADTETAPDVDVRTAARPRTAAR 1400

Qy      58 KLPPADSADGQAAVDVHNAQITALIETRASRLHFEG--ETPATIA----DTFAKAEKLDR 111
```

```
Db   1401 TSPRVDTRTGARTAD--GPRLVVLTRGAAGP---EGGAADPAGAAVWGLVRVAQAEQPGR 1455

Qy    112 LATT----TSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLI---------- 155

Db   1456 FTLVDVDGTQASLRALPGLLAT--DAGQSAVRDGRVTVPRLVPVADPVPHGGGTAADGTG 1513

Qy    156 SGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRH---------------- 199

Db   1514 AGEPSATLDPEGTVLITGGTG-------------ALAAETARHLVDRHKVRHLLLVGRR 1559

Qy    200 ---SPSLARQVLD-----TGVAVQT--YSARNAVRTVLAPALASRPAVQGAVDLGVSMAG 249

Db   1560 GPDAPGVDRLVAELTESGAEVAVRACDVTDRDALRRLLDALPDEHPLTCVVHTAGVLDDG 1619

Qy    250 GLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLE---------AYKA 300

Db   1620 VLSAQT--AERIDTV-----------------LRPKADAAVHLDELTREIGRVPLVLYSS 1660

Qy    301 IKS-------ASYSGA-----ALNAGKRMAGLP-----------LDMATDAMGAVRSLVS 337

Db   1661 VSATLGSAGQAGYAAANAFMDALAARRCAAGHPALSLGWGWWSGVGLATGLDGA-----D 1715

Qy    338 ASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAV------FAGW 391

Db   1716 AARVRRSGLAPLDAGAALDLLDRALTR--PEPALLPVRLDLRAAAGATALPEVLRDLAGV 1773

Qy    392 TTAALTT----------DPAVKKA---------------------ESFIQDTVKSTASS 419

Db   1774 PADARSTPGAAAGTGDEDGAVRPAPAPADAAGTLAARLAGRSAPERTALLLDLVRTEVAA 1833

Qy    420 TTGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVN---NLR---------QRPAREAD- 466

Db   1834 VLGHGDPAAIGAARTFKDAGFDSL--TAVDLRNRLNTRTGLRLPATLVFDHPTPLALAEL 1891

Qy    467 ----IEEGGTAASPSEIP 480

Db   1892 LLDGLEAAGPAEPAAEVP 1909


RESULT 2
US-08-804-198-3
; Sequence 3, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1996 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-3

  Query Match             5.4%;  Score 130.5;  DB 2;  Length 1996;
  Best Local Similarity   21.5%;  Pred. No. 0.008;
  Matches  133;  Conservative   70;  Mismatches  216;  Indels  199;  Gaps   28;

Qy     12 VTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHV-------------FAAH 57

Db   1342 LTTTAAGRTASDGG-RLSLLDEVRATVRRTLEAVQARLADTETAPDVDVRTAARPRTAAR 1400

Qy     58 KLPPADSADGQAAVDVHNAQITALIETRASRLHFEG--ETPATIA----DTFAKAEKLDR 111

Db   1401 TSPRVDTRTGARTAD--GPRLVVLTRGAAGP---EGGAADPAGAAVWGLVRVAQAEQPGR 1455

Qy    112 LATT----TSGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLI---------- 155

Db   1456 FTLVDVDGTQASLRALPGLLAT--DAGQSAVRDGRVTVPRLVPVADPVPHGGGTAADGTG 1513

Qy    156 SGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRH---------------- 199

Db   1514 AGEPSATLDPEGTVLITGGTG-------------ALAAETARHLVDRHKVRHLLLVGRR 1559

Qy    200 ---SPSLARQVLD-----TGVAVQT--YSARNAVRTVLAPALASRPAVQGAVDLGVSMAG 249

Db   1560 GPDAPGVDRLVAELTESGAEVAVRACDVTDRDALRRLLDALPDEHPLTCVVHTAGVLDDG 1619

Qy    250 GLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLE---------AYKA 300

Db   1620 VLSAQT--AERIDTV-----------------LRPKADAAVHLDELTREIGRVPLVLYSS 1660

Qy    301 IKS-------ASYSGA-----ALNAGKRMAGLP-----------LDMATDAMGAVRSLVS 337

Db   1661 VSATLGSAGQAGYAAANAFMDALAARRCAAGHPALSLGWGWWSGVGLATGLDGA-----D 1715

Qy    338 ASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAV------FAGW 391

Db   1716 AARVRRSGLAPLDAGAALDLLDRALTR--PEPALLPVRLDLRAAAGATALPEVLRDLAGV 1773

Qy    392 TTAALTT----------DPAVKKA---------------------ESFIQDTVKSTASS 419

Db   1774 PADARSTPGAAAGTGDEDGAVRPAPAPADAAGTLAARLAGRSAPERTALLLDLVRTEVAA 1833

Qy    420 TTGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVN---NLR---------QRPAREAD- 466

Db   1834 VLGHGDPAAIGAARTFKDAGFDSL--TAVDLRNRLNTRTGLRLPATLVFDHPTPLALAEL 1891

Qy    467 ----IEEGGTAASPSEIP 480

Db   1892 LLDGLEAAGPAEPAAEVP 1909


RESULT 3
US-08-194-290-7
; Sequence 7, Application US/08194290
; Patent No. 5500353
;  GENERAL INFORMATION:
;    APPLICANT:  Smit, John
;    APPLICANT:  Bingle, Wade H
;    TITLE OF INVENTION:  Bacterial surface protein expression
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shlesinger, Arkwright & Garvey
;      STREET:  3000 South Eads Street
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
```

```
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,290
;      FILING DATE:  09-FEB-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garvey, George A
;      REGISTRATION NUMBER:  17737
;      REFERENCE/DOCKET NUMBER:  5946-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-684-5600
;      TELEFAX:  703-836-5288
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1026 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-194-290-7

  Query Match             5.3%;  Score 128.5;  DB 1;  Length 1026;
  Best Local Similarity   22.4%;  Pred. No. 0.0043;
  Matches  114;  Conservative   57;  Mismatches  175;  Indels  163;  Gaps   24;

Qy      12 VTATDSFRTASDASLASSSVRSVSSDQQREINAIADY---------------------- 48
           |:    : ||: | :   |  ::|:|    :|    |
Db     206 VSGIGGYATATAAMINDLSDGALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTA 265

Qy      49 -----LTDHVFAAHKLPPADSADGQAAVDVHN----AQITAL--------IETRASRLHF 91
                :   |  |  |   |:  | |  || |    | :|||        |||    ::
Db     266 NNDTFVAGEVAGAATLTVGDTLSGGAGTDVLNWVQAAAVTALPTGVTISGIET----MNV 321

Qy      92 EGETPATIADTFAKAEKLDRLATTTSGA----------LRATPFAMASLLQYMQPAINK 140
              |: :| :    |  | | ||||           | ||  | |:       |::
Db     322 TSGAAITL-NTSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAA----NNVAVDG 376

Qy     141 GDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHS 200
           |         :|   :|  ||    ||      |:|                :: ||   |
Db     377 G-------ANVTVASTGVTSGT-TTVGAN--SAASG-------------TVSVSVANSS 412

Qy     201 PSLARQVLDTGVAVQT--YSARNAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFG 258
            :    : ||     |   :| ||| | |  |  :         : |:        | |
Db     413 TTTTGAIAVTGGTAVTVAQTAGNAVNTTLTQADVTVTGNSSTTAVTVTQTAAATAGATVA 472

Qy     259 NRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMA 318
           |:           ||: :                    : | : :|  |||
Db     473 GRV----------NGAVTI-----------------------TDSAAASATTAGK--- 494

Qy     319 GLPLDMATDAMGAV-RSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQ 377
           :||  :|:   : : :|:||   |:  |   |:|:        :|  ||  | :
Db     495 -----IATVTLGSFGAATIDSSALTTVNLSGTGTSLGIGR------GALT--ATPTANTL 541

Qy     378 LTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQD---TVK---STASSTTGYVADQTVKL 431
             |: |           |||  |:  :|:   |   |:    ||||||   :|
Db     542 TLNVNG-----------LTTTGAITDSEAAADDGFTTINIAGSTASST---IASLVAAD 586

Qy     432 AKTVKDMGGEAIT---HTGASLRN-TVNN 456
           | |:   |   :|   || |:|   || |
Db     587 ATTLNISGDARVTITSHTAAALTGITVTN 615


RESULT 4
US-08-591-079-6
; Sequence 6, Application US/08591079
; Patent No. 5972899
;  GENERAL INFORMATION:
;    APPLICANT:  Zychlinksky, Arturo
;    APPLICANT:  Chen, Yajing
;    TITLE OF INVENTION:  Apoptosis Induced by Shigella IpaB
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,079
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Livnat, Shmuel
;      REGISTRATION NUMBER:  33,949
;      REFERENCE/DOCKET NUMBER:  15661-20017.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0764
;      TELEX:  90-4030 MRSNFOERSWSH
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 401 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-591-079-6

  Query Match             5.3%;  Score 127.5;  DB 2;  Length 401;
  Best Local Similarity   19.7%;  Pred. No. 0.0012;
  Matches   82;  Conservative   57;  Mismatches  124;  Indels  153;  Gaps   17;

Qy     121 RATPFAMASLLQYMQ---PA-----------------INKGDWLPAPLKPLTPLISGAL 159
           |:||    |||  |::   ||                    ::|  |||||     :::. :
Db       9 RSTP-VTGSLLPYVETPAPAPLQTQQVAGELKDKNGGVSSQGVQLPAPL----AVVASQV 63

Qy     160 SGAMDQVGTKMMDRATGDL-------HYLS-------ASPD------------------- 186
           :    |  ||:::  |           :|:|       ||||
Db      64 TEGQQQEVTKLLESVTRGAAGSQLISNYVSVLTKFTLASPDTFEIELGKLVSNLEEVRKD 123

Qy     187 -------RLHDAMAASVKRHSPSL------ARQVLDTGVAVQTYSARNAVRTVLAPALAS 233
                  |||:      :: :   :       |:||  :|:|  : :   :|: :|:  |:
Db     124 IKIADIQRLHEQNMKKIEENQEKIKETEENAKQVKKSGIASKIFGWLSAIASVIVGAIMV 183

Qy     234 RPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEEND 293
           | |||       || :   :|                    |:|:
Db     184 ASGV-GAV------AGAMMVASG------------------VIGM-------------- 203

Qy     294 WLEAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFA 353
           |  |:| |:  |        :: |  |      |:  | ::::         ||  |
Db     204 ---ANMAVKQAAEDGLISQEAMKILGPILTAIEVALTVVSTVMTFGGSALKCLANIGAKL 260

Qy     354 GVGKLQEMATKNITDPATKAAVSQ-LTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDT 412
           |      :| |      |  | :|  ::|  |||     | : | ||    |::
Db     261 G-ANTASLAAKGAEFSAKVAQISTGISNTVGSAVTKLGGSFAGLTMSHAIR--------- 310

Qy     413 VKSTASSTTGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIE 468
           | |  |                    :      |: :  |:|| :|      ||::
Db     311 ---TGSQAT-----------------QVAVGVGSGITQTINNKKQ-----ADLQ 339


RESULT 5
US-09-085-199B-7
; Sequence 7, Application US/09085199B
; Patent No. 6235879
;  GENERAL INFORMATION:
;    APPLICANT:  Hayden. Michael R.
```

```
;    APPLICANT:  Hackam, Abigail
;    APPLICANT:  Huq, A.H.M. Mahbubul
;    APPLICANT:  Chopra, Vikramjit Singh
;    APPLICANT:  Kalchman, Michael
;    TITLE OF INVENTION:  Apoptosis Modulators That Interact with the
;    TITLE OF INVENTION:  Huntington's Disease Gene
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oppedahl & Larson
;      STREET:  PO Box 5270
;      CITY:  Frisco
;      STATE:  CO
;      COUNTRY:  USA
;      ZIP:  80443-5270
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Kb storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS DOS 5.0
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/085,199B
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Larson, Marina T.
;      REGISTRATION NUMBER:  32038
;      REFERENCE/DOCKET NUMBER:  UBC.P-013US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (970) 668-2050
;      TELEFAX:  (970) 668-2052
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  676
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  human
;    FEATURE:
;      OTHER INFORMATION:  Huntington-interacting protein
US-09-085-199B-7

  Query Match              5.3%;  Score 127;  DB 4;  Length 676;
  Best Local Similarity   21.4%;  Pred. No. 0.003;
  Matches  117;  Conservative   68;  Mismatches  209;  Indels  154;  Gaps   24;

Qy      33 SVSSDQQREINAIADYLT---DHVFAAHKLPPADSADGQAAVDVHNAQITALIETRASRL 89
           | :   : |:::  | |:   | :  | :   ||    |: |    |   ||   :
Db     144 SHTEQSKSELSSRLDTLSAEKDALSGAVRQREADLLAAQSLVRETEA---ALSREQQRSS 200

Qy      90 HFEGETPATIADTFAKAEKL-DRLATTTSGALRATPFAMASLLQYMQPAINKGDWLPAPL 148
             :||    :|:  :: : |  ||        ||       | :|   | |::| |
Db     201 QEQGELQGRLAERESQEQGLRQRLLDEQFAVLRGAAAEAAGIL---QDAVSKLD------ 251

Qy     149 KPL------TP--LISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHS 200
           ||       :|  |:| |    |:| | |       |   ||::  |    |: |:: | |
Db     252 DPLHLRCTSSPDYLVSRA-QEALDAVST----LEEGHAQYLTSLADA--SALVAALTRFS 304

Qy     201 PSLARQVLDTGVAVQTYSARNAVRTV----------------------LAPALAS--RPA 236
             |  ::: |           | | :                     |    ||  |
Db     305 HLAADTIINGGATSHLAPTDPADRLIDTCRECGARALELMGQLQDQQALRHMQASLVRTP 364

Qy     237 VQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLE 296
           :|| : ||  :           :  |:  |  : |  :     ||| |  :   | :
Db     365 LQGILQLGQEL-----------KPKSLDVRQEELGAVV-----DKEMAATSAAIEDAVR 407

Qy     297 AYKAIKSA---SYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFA 353
             : : :    : ||  |   :|:    |:  || | |:| ||: |:  |  :  :|
Db     408 RIEDMMNQARHASSGVKLEVNERI----LNSCTDLMKAIRLLVTTSTSLQKEIVESG--R 461

Qy     354 GVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTV 413
           |    ||   ||          |: |    ||:   ||    |     |: |:  :  |
```

```
Db     462 GAATQQEFYAKN----------SRWTEGLISASKAVGWGATQL-----VEAADKVVLHTG 506

Qy     414 K---------STASSTTGYVADQTVK-------------LAKTVKDMGGEAITHT----- 446
           |           |:||   ||   ||               ::|| :     :  |
Db     507 KYEELIVCSHEIAASTAQLVAASKVKANKHSPHLSRLQECSRTVNERAANVVASTKSGQE 566

Qy     447 ----------GASLRNTVNNLRQRPAREADIEE------------------GGTAASP 476
                    | ||        :   |  ::|:                  | : ||
Db     567 QIEDRDTMDFSGLSLIKLKKQEMETQVRVLELEKTLEAERMRLGELRKQHYVLAGASGSP 626

Qy     477 -SEIPFRP 483
            |:  ||
Db     627 GEEVAIRP 634


RESULT 6
US-08-968-685A-10
; Sequence 10, Application US/08968685A
; Patent No. 6214981
;  GENERAL INFORMATION:
;    APPLICANT:  TUCKER, KENNETH
;    APPLICANT:  PLOSILA, LAURA
;    TITLE OF INVENTION:  MORAXELLA CATARRHALIS OUTER MEMBRANE
;    TITLE OF INVENTION:  PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/968,685A
;      FILING DATE:  No. 6214981ember 12, 1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baldwin, Geraldine F.
;      REGISTRATION NUMBER:  31,232
;      REFERENCE/DOCKET NUMBER:  7969-060
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2123 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-968-685A-10

  Query Match              5.3%;  Score 127;  DB 4;  Length 2123;
  Best Local Similarity   21.1%;  Pred. No. 0.019;
  Matches  116;  Conservative   84;  Mismatches  218;  Indels  132;  Gaps   27;

Qy      23 DASLASSSVRSVSSDQQREIN-----AIADYLTDHVFAAHKLPPADS--------ADGQA 69
           |  : :::||: : |  : |:         :  || ||:          ||         | |
Db     281 DNDVNAANVRAYAPDDNQPIDNRYKATFKNGATD-VFSIGNSNGNDSIRRKIINVGAGSA 339

Qy      70 AVDVHN-AQITALIETRASRLHFEGETPATIADTFAKAEK-LDRLATTTSGALRATPFAM 127
             |  | ||:   :      :: |:|:       |:  : || | :  | | ||
Db     340 DTDAVNVAQLKEAVRLANRQITFKGD------DSNNRVEKGLGKTLTITGGA------QT 387

Qy     128 ASLLQYMQPAINKGDWLPAPL-KPLTPL--------------ISGALSGAMDQVG----- 167
           ::|  :    :  || |   | : || |                |      |::|
Db     388 SALTDHNIGVVQNGDGLKVQLAETLTSLKMVTTENLTANEKVTVGKTRLTTDKIGFTNDM 447
```

```
Qy    168 ------TKMMDRATGDLHYLSASPDRLHDAMAASVKRHSPSLA--RQVLDTG-VAVQTYS 218
                  :|: ||    : |   ::    |  |   : :    ::|  |   |:||::
Db    448 NGIDESKPYLDKDTG----IHAGGQKITKLTAGVVDDDAATYGQLKKVNQTAESALQTFT 503

Qy    219 ARNAVRT--------VLAPALASRP-------------------AVQGAVDLGVSMAGG 250
           :   :          ::     ::|                   | |  |::   |
Db    504 VKKVDKNGNDANDSKIITVGKNNKPDGTQVNTLKLKGENGVDVTTETNGTVTFGLNQNNG 563

Qy    251 L-AANAGFGNRLLSVQSRDHQR------GGALVLGLKDKEPKAQLSEENDWLEAYKAIKS 303
           |   |:   |  |||:: :  :       |     : : :| | :  ||        |
Db    564 LTVGNSTLNNDGLSVKNTNSNKQIQVGADGITFTDISNSKPGAGI--ENTTRITRDGIGF 621

Qy    304 ASYSGAALNAGK-RMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKL---Q 359
           |: :| :|:|  | |:   ::     :  |:| :: ::   ||          |:|
Db    622 ANNTG-SLDANKPRLTPTGINAGGKELTNVQSAINPAT---NG----------GQLDFMN 667

Qy    360 EMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASS 419
           ::| |     : | :  | ||:     ||| |   :|     | |       |:||
Db    668 RLSTANTEKSGSAATIKDLYNLSQVPLTFAGDTGPNVTK----KLGEILKVKGGKTTADD 723

Qy    420 TT----GYVADQ-----TVKLAKTVKDMGGEAITHTGASLRNTV---NNLRQRPAREADI 467
           |     | |||      |||||||: |:          || : ||   ||  :    :
Db    724 LTKNNIGVVADSTDNSLTVKLAKTLSDLDAVNTKTLTASDKVTVDSGNNTAKLQNGDLTF 783

Qy    468 EEGGTAASPS 477
           :  | |:|:
Db    784 SKQNTGATPA 793
```

RESULT 7
US-09-071-035-264
; Sequence 264, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 264:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 669 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-264

Query Match 5.2%; Score 126.5; DB 4; Length 669;
Best Local Similarity 20.8%; Pred. No. 0.0033;
Matches 113; Conservative 69; Mismatches 192; Indels 169; Gaps 27;

```
Qy     17 SFRT-ASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHN 75
          :|:| | :|:: : ::  |:       |: | |                 :||  | |:: |
Db    116 TFQTPADEATVTAPTITGVTG------NSTAGY----------EVKGTADANATVEIRN 158

Qy     76 AQITALIETRASRLHFEGETPATIADTFAKA---EKLDRLATTTSGALRATPFAMASLLQ 132
          |   |:| |  :    :|    |:     :|   | |  :|   ||   :||    :
Db    159 AG-GAVIGTGTA----DGTGAFTVTIPAGEAGANETLTAVAKNASGT-ESTPTTFQT--- 209

Qy    133 YMQPAINKGDWLPAPLKPLTPL-------ISGALSGAMD-----QVGT--KMMDRATGDL 178
                      | | ||:       ::|: :   :     :|||  :: | |   |
Db    210 --------------PADPNTPVATPIVETVTGSTTKGYEVKGTAEVGTTIEVRDAAGTVL 255

Qy    179 HYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAV--QTYS--ARNAVRTVLAPALASR 234
             :   | :                    ||:| |   || |  |:||  |   || |:
Db    256 GTATTGTDGKYTV--------------TLDSGTATANQTLSVVAKNASGTESQPATATT 300

Qy    235 PAVQGAVDLGVSMAGGLAANAGFGNRL-------LSVQSRDHQRGGALVLGLKDKEPKAQ 287
          ||      |:      :  |:| |  :         ::: ||   |  :  |  |
Db    301 PA-----DVTAPTVDNITGNSGSGYEITGTADPNTTIEVRD-PSGAVIGTGTSD------ 348

Qy    288 LSEENDWLEAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSA--------- 338
             |:       : :      | |  :  :  | | :       || |
Db    349 --ANGDF--------TVTLPTGTTNPGDTLTVIGKDNAGNESQPTEVLVPADATVTAPTV 398

Qy    339 SSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAV-------FAGW 391
          : :| | :|   |:    |     || | |       :   :  || ||         |
Db    399 TGVTGNSVA---GYQVTGTADPNATIEIRDADGNVIATGTADGTGSFAVNLPAGTANANE 455

Qy    392 TTAALTTDPAVKKAESFIQDTV----------KSTASSTTGYVADQTVKLAKTVK---- 436
          |  ||  |||   :      |             | | ::| ||    | :|  |::
Db    456 TLTALAKDPAGNTSTPTTFQTPADEVVAPPSVDKVTGNTTQGYQVTGTAELGTTIEVRAT 515

Qy    437 ---DMGGEAITHTG----------ASLRNTVN-------NLRQRPAREADIEEGGTAASP 476
             :|      ||          |: : |||        |  :|          || :|
Db    516 DGTVLGTATTGPTGQYTVTLASGKATAKQTVNVVAKNDTGLESQPT---------TAMTP 566

Qy    477 SEI 479
             :::
Db    567 ADV 569
```

RESULT 8
US-08-614-377A-7
; Sequence 7, Application US/08614377A
; Patent No. 5976864
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 5976864ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
; TITLE OF INVENTION: HETEROLOGOUS
; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson PC
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/614,377A
; FILING DATE: 12-MAR-1996
; CLASSIFICATION: 435

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/194,290
;      FILING DATE:  09-FEB-1994
;      CLASSIFICATION:  435
;      APPLICATION NUMBER:  US 07/895,367
;      FILING DATE:  09-JUNE-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Tsao, Y. Rocky
;      REGISTRATION NUMBER:  34053
;      REFERENCE/DOCKET NUMBER:  08106/002001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-542-5070
;      TELEFAX:  617-542-8906
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1026 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-614-377A-7

  Query Match             5.2%;  Score 126.5;  DB 2;  Length 1026;
  Best Local Similarity   22.2%;  Pred. No. 0.0066;
  Matches  113;  Conservative   56;  Mismatches  177;  Indels  163;  Gaps   23;

Qy      12 VTATDSFRTASDASLASSSVRSVSSDQQREINAIADY---------------------- 48
           |:    : ||: | :   |  ::|:|    :|    |
Db     206 VSGIGGYATATAAMINDLSDGALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTA 265

Qy      49 -----LTDHVFAAHKLPPADSADGQAAVDVHN----AQITAL--------IETRASRLHF 91
                :   |  |  |   |:  | |  || |    | :|||         |||    ::
Db     266 NNDTFVAGEVAGAATLTVGDTLSGGAGTDVLNWVQAAAVTALPTGVTISGIET----MNV 321

Qy      92 EGETPATIADTFAKAEKLDRLATTTSGA----------LRATPFAMASLLQYMQPAINK 140
              |: :| :    |  | | ||||           | ||  | |:    :    |
Db     322 TSGAAITL-NTSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGRAN- 379

Qy     141 GDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHS 200
                  :|   :|  ||     ||      |:|             :: ||   |
Db     380 ----------VTVASTGVTSGT-TTVGAN--SAASG-------------TVSVSVANSS 412

Qy     201 PSLARQVLDTGVAVQT--YSARNAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFG 258
           :    :  ||    |   :| |||  | |  |  :         : |:      | |
Db     413 TTTTGAIAVTGGTAVTVAQTAGNAVNTTLTQADVTVTGNSSTTAVTVTQTAAATAGATVA 472

Qy     259 NRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMA 318
           |:          ||: :                  : | : :|  |||
Db     473 GRV----------NGAVTI-------------------------TDSAAASATTAGK--- 494

Qy     319 GLPLDMATDAMGAV-RSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQ 377
              :||  :|:   : : :|:||   |:  |   |:|:        :|  ||  | :
Db     495 -----IATVTLGSFGAATIDSSALTTVNLSGTGTSLGIGR------GALT--ATPTANTL 541

Qy     378 LTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQD---TVK---STASSTTGYVADQTVKL 431
             |: |          |||  |:  :|:    |    |:    ||||||   :|
Db     542 TLNVNG----------LTTTGAITDSEAAADDGFTTINIAGSTASST---IASLVAAD 586

Qy     432 AKTVKDMGGEAIT---HTGASLRN-TVNN 456
           | |:    |   :|    || |:|    || |
Db     587 ATTLNISGDARVTITSHTAAALTGITVTN 615


RESULT 9
US-09-142-648B-7
; Sequence 7, Application US/09142648B
; Patent No. 6210948
; GENERAL INFORMATION:
;  APPLICANT:  Smit, John
;  APPLICANT:  Bingle, Wade H.
;  APPLICANT:  No. 6210948ellini, John F.
;  TITLE OF INVENTION:  EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
;  FILE REFERENCE:  08106/002002
;  CURRENT APPLICATION NUMBER: US/09/142,648B
;  CURRENT FILING DATE:  1999-03-30
;  PRIOR APPLICATION NUMBER:  PCT/CA97/00167
;  PRIOR FILING DATE:  1997-03-10
;  PRIOR APPLICATION NUMBER:  US 07/614,377
;  PRIOR FILING DATE:  1996-03-12
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 1026
;   TYPE: PRT
;   ORGANISM: Caulobacter crescentus
US-09-142-648B-7

  Query Match             5.2%;  Score 126.5;  DB 4;  Length 1026;
  Best Local Similarity   22.2%;  Pred. No. 0.0066;
  Matches  113;  Conservative   56;  Mismatches  177;  Indels  163;  Gaps   23;

Qy      12 VTATDSFRTASDASLASSSVRSVSSDQQREINAIADY---------------------- 48
           |:    : ||: | :   |  ::|:|    :|    |
Db     206 VSGIGGYATATAAMINDLSDGALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTA 265

Qy      49 -----LTDHVFAAHKLPPADSADGQAAVDVHN----AQITAL--------IETRASRLHF 91
                :   |  |  |   |:  | |  || |    | :|||         |||    ::
Db     266 NNDTFVAGEVAGAATLTVGDTLSGGAGTDVLNWVQAAAVTALPTGVTISGIET----MNV 321

Qy      92 EGETPATIADTFAKAEKLDRLATTTSGA----------LRATPFAMASLLQYMQPAINK 140
              |: :| :    |  | | ||||           | ||  | |:    :    |
Db     322 TSGAAITL-NTSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGRAN- 379

Qy     141 GDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHYLSASPDRLHDAMAASVKRHS 200
                  :|   :|  ||     ||      |:|             :: ||   |
Db     380 ----------VTVASTGVTSGT-TTVGAN--SAASG-------------TVSVSVANSS 412

Qy     201 PSLARQVLDTGVAVQT--YSARNAVRTVLAPALASRPAVQGAVDLGVSMAGGLAANAGFG 258
           :    :  ||    |   :| |||  | |  |  :         : |:      | |
Db     413 TTTTGAIAVTGGTAVTVAQTAGNAVNTTLTQADVTVTGNSSTTAVTVTQTAAATAGATVA 472

Qy     259 NRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKAIKSASYSGAALNAGKRMA 318
           |:          ||: :                  : | : :|  |||
Db     473 GRV----------NGAVTI-------------------------TDSAAASATTAGK--- 494

Qy     319 GLPLDMATDAMGAV-RSLVSASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQ 377
              :||  :|:   : : :|:||   |:  |   |:|:        :|  ||  | :
Db     495 -----IATVTLGSFGAATIDSSALTTVNLSGTGTSLGIGR------GALT--ATPTANTL 541

Qy     378 LTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQD---TVK---STASSTTGYVADQTVKL 431
             |: |          |||  |:  :|:    |    |:    ||||||   :|
Db     542 TLNVNG----------LTTTGAITDSEAAADDGFTTINIAGSTASST---IASLVAAD 586

Qy     432 AKTVKDMGGEAIT---HTGASLRN-TVNN 456
           | |:    |   :|    || |:|    || |
Db     587 ATTLNISGDARVTITSHTAAALTGITVTN 615


RESULT 10
US-09-071-035-258
; Sequence 258, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
```

```
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  258:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1638 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-258

  Query Match             5.2%;  Score 126.5;  DB 4;  Length 1638;
  Best Local Similarity   20.8%;  Pred. No. 0.014;
  Matches  113;  Conservative   69;  Mismatches  192;  Indels  169;  Gaps   27;

Qy      17 SFRT-ASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHN 75
           :|:| | :|:: : ::  |:       |: | ||              :||  | |:: |
Db     523 TFQTPADEATVTAPTITGVTG------NSTAGY----------EVKGTADANATVEIRN 565

Qy      76 AQITALIETRASRLHFEGETPATIADTFAKA---EKLDRLATTTSGALRATPFAMASLLQ 132
           |   |:| |  :    :|    |:     :|   | |  :|   ||   :||    :
Db     566 AG-GAVIGTGTA----DGTGAFTVTIPAGEAGANETLTAVAKNASGT-ESTPTTFQT--- 616

Qy     133 YMQPAINKGDWLPAPLKPLTPL-------ISGALSGAMD-----QVGT--KMMDRATGDL 178
                   | | ||:       ::|: :   :     :|||  :: | |   |
Db     617 --------------PADPNTPVATPIVETVTGSTTKGYEVKGTAEVGTTIEVRDAAGTVL 662

Qy     179 HYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAV--QTYS--ARNAVRTVLAPALASR 234
              :   |  :                ||:| |   || |  |:||  |   || |:
Db     663 GTATTGTDGKYTV--------------TLDSGTATANQTLSVVAKNASGTESQPATATT 707

Qy     235 PAVQGAVDLGVSMAGGLAANAGFGNRL-------LSVQSRDHQRGGALVLGLKDKEPKAQ 287
           ||     |:       :  |:| |  :        ::: ||   |  :  |  |
Db     708 PA-----DVTAPTVDNITGNSGSGYEITGTADPNTTIEVRD-PSGAVIGTGTSD------ 755

Qy     288 LSEENDWLEAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSA--------- 338
                 |:        : :      | |  :  :  | | :       || |
Db     756 --ANGDF--------TVTLPTGTTNPGDTLTVIGKDNAGNESQPTEVLVPADATVTAPTV 805

Qy     339 SSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAV-------FAGW 391
           : :| | :|    |:   |      ||  | |        :   :  || ||        |
Db     806 TGVTGNSVA---GYQVTGTADPNATIEIRDADGNVIATGTADGTGSFAVNLPAGTANANE 862

Qy     392 TTAALTTDPAVKKAESFIQDTV----------KSTASSTTGYVADQTVKLAKTVK---- 436
           |  ||  |||   :       |              | | ::| ||    | :|  |::
Db     863 TLTALAKDPAGNTSTPTTFQTPADEVVAPPSVDKVTGNTTQGYQVTGTAELGTTIEVRAT 922

Qy     437 ---DMGGEAITHTG----------ASLRNTVN-------NLRQRPAREADIEEGGTAASP 476
              :|      ||           |: : |||         | :|           || :|
Db     923 DGTVLGTATTGPTGQYTVTLASGKATAKQTVNVVAKNDTGLESQPT---------TAMTP 973

Qy     477 SEI 479
           :::
Db     974 ADV 976


RESULT 11
US-09-071-035-262
```

```
; Sequence 262, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette; 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  262:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1638 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-262

  Query Match             5.2%;  Score 126.5;  DB 4;  Length 1638;
  Best Local Similarity   20.8%;  Pred. No. 0.014;
  Matches  113;  Conservative   69;  Mismatches  192;  Indels  169;  Gaps   27;

Qy      17 SFRT-ASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHN 75
           :|:| | :|:: : ::  |:       |: | ||              :||  | |:: |
Db     523 TFQTPADEATVTAPTITGVTG------NSTAGY----------EVKGTADANATVEIRN 565

Qy      76 AQITALIETRASRLHFEGETPATIADTFAKA---EKLDRLATTTSGALRATPFAMASLLQ 132
           |   |:| |  :    :|    |:     :|   | |  :|   ||   :||    :
Db     566 AG-GAVIGTGTA----DGTGAFTVTIPAGEAGANETLTAVAKNASGT-ESTPTTFQT--- 616

Qy     133 YMQPAINKGDWLPAPLKPLTPL-------ISGALSGAMD-----QVGT--KMMDRATGDL 178
                   | | ||:       ::|: :   :     :|||  :: | |   |
Db     617 --------------PADPNTPVATPIVETVTGSTTKGYEVKGTAEVGTTIEVRDAAGTVL 662

Qy     179 HYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAV--QTYS--ARNAVRTVLAPALASR 234
              :   |  :                ||:| |   || |  |:||  |   || |:
Db     663 GTATTGTDGKYTV--------------TLDSGTATANQTLSVVAKNASGTESQPATATT 707

Qy     235 PAVQGAVDLGVSMAGGLAANAGFGNRL-------LSVQSRDHQRGGALVLGLKDKEPKAQ 287
           ||     |:       :  |:| |  :        ::: ||   |  :  |  |
Db     708 PA-----DVTAPTVDNITGNSGSGYEITGTADPNTTIEVRD-PSGAVIGTGTSD------ 755

Qy     288 LSEENDWLEAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSA--------- 338
                 |:        : :      | |  :  :  | | :       || |
Db     756 --ANGDF--------TVTLPTGTTNPGDTLTVIGKDNAGNESQPTEVLVPADATVTAPTV 805

Qy     339 SSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAV-------FAGW 391
           : :| | :|    |:   |      ||  | |        :   :  || ||        |
Db     806 TGVTGNSVA---GYQVTGTADPNATIEIRDADGNVIATGTADGTGSFAVNLPAGTANANE 862
```

```
Qy     392 TTAALTTDPAVKKAESFIQDTV-----------KSTASSTTGYVADQTVKLAKTVK---- 436
           |  ||  |||   :      |             | | ::| ||    | :|  |::
Db     863 TLTALAKDPAGNTSTPTTFQTPADEVVAPPSVDKVTGNTTQGYQVTGTAELGTTIEVRAT 922

Qy     437 ---DMGGEAITHTG----------ASLRNTVN-------NLRQRPAREADIEEGGTAASP 476
               :|       ||          |: : |||        |  :|           || :|
Db     923 DGTVLGTATTGPTGQYTVTLASGKATAKQTVNVVAKNDTGLESQPT---------TAMTP 973

Qy     477 SEI 479
           :::
Db     974 ADV 976


RESULT 12
US-09-071-035-266
; Sequence 266, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  266:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1638 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-266

  Query Match             5.2%;  Score 126.5;  DB 4;  Length 1638;
  Best Local Similarity   20.8%;  Pred. No. 0.014;
  Matches  113;  Conservative   69;  Mismatches  192;  Indels  169;  Gaps   27;

Qy      17 SFRT-ASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHN 75
           :|:| | :|:: : ::  |:       |: | |              :||  | |:: |
Db     523 TFQTPADEATVTAPTITGVTG------NSTAGY----------EVKGTADANATVEIRN 565

Qy      76 AQITALIETRASRLHFEGETPATIADTFAKA---EKLDRLATTTSGALRATPFAMASLLQ 132
           |   |:| |  :    :|    |:     :|   | |  :|   ||   :||    :
Db     566 AG-GAVIGTGTA----DGTGAFTVTIPAGEAGANETLTAVAKNASGT-ESTPTTFQT--- 616

Qy     133 YMQPAINKGDWLPAPLKPLTPL-------ISGALSGAMD-----QVGT--KMMDRATGDL 178
                     | | ||:       ::|: :   :     :|||  :: | |   |
Db     617 -------------PADPNTPVATPIVETVTGSTTKGYEVKGTAEVGTTIEVRDAAGTVL 662

Qy     179 HYLSASPDRLHDAMAASVKRHSPSLARQVLDTGVAV--QTYS--ARNAVRTVLAPALASR 234
           :::
Db     663 GTATTGTDGKYTV---------------TLDSGTATANQTLSVVAKNASGTESQPATATT 707

Qy     235 PAVQGAVDLGVSMAGGLAANAGFGNRL-------LSVQSRDHQRGGALVLGLKDKEPKAQ 287
           ||     |:       :  |:| | :         ::: ||   |  :  |  |
Db     708 PA-----DVTAPTVDNITGNSGSGYEITGTADPNTTIEVRD-PSGAVIGTGTSD------ 755

Qy     288 LSEENDWLEAYKAIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSA--------- 338
               |:        : :      | |  :  :  | | :       || |
Db     756 --ANGDF--------TVTLPTGTTNPGDTLTVIGKDNAGNESQPTEVLVPADATVTAPTV 805

Qy     339 SSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAV-------FAGW 391
           : :| | :|   |:   |      ||  | |       :   :  || ||       |
Db     806 TGVTGNSVA---GYQVTGTADPNATIEIRDADGNVIATGTADGTGSFAVNLPAGTANANE 862

Qy     392 TTAALTTDPAVKKAESFIQDTV-----------KSTASSTTGYVADQTVKLAKTVK---- 436
           |  ||  |||   :      |             | | ::| ||    | :|  |::
Db     863 TLTALAKDPAGNTSTPTTFQTPADEVVAPPSVDKVTGNTTQGYQVTGTAELGTTIEVRAT 922

Qy     437 ---DMGGEAITHTG----------ASLRNTVN-------NLRQRPAREADIEEGGTAASP 476
               :|       ||          |: : |||        |  :|           || :|
Db     923 DGTVLGTATTGPTGQYTVTLASGKATAKQTVNVVAKNDTGLESQPT---------TAMTP 973

Qy     477 SEI 479
           :::
Db     974 ADV 976


RESULT 13
US-09-085-199B-11
; Sequence 11, Application US/09085199B
; Patent No. 6235879
;  GENERAL INFORMATION:
;    APPLICANT:  Hayden. Michael R.
;    APPLICANT:  Hackam, Abigail
;    APPLICANT:  Huq, A.H.M. Mahbubul
;    APPLICANT:  Chopra, Vikramjit Singh
;    APPLICANT:  Kalchman, Michael
;    TITLE OF INVENTION:  Apoptosis Modulators That Interact with the
;    TITLE OF INVENTION:  Huntington's Disease Gene
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oppedahl & Larson
;      STREET:  PO Box 5270
;      CITY:  Frisco
;      STATE:  CO
;      COUNTRY:  USA
;      ZIP:  80443-5270
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Kb storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS DOS 5.0
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/085,199B
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Larson, Marina T.
;      REGISTRATION NUMBER:  32038
;      REFERENCE/DOCKET NUMBER:  UBC.P-013US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (970) 668-2050
;      TELEFAX:  (970) 668-2052
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1068
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  mouse
```

```
;    FEATURE:
;      OTHER INFORMATION:  Huntington-interacting protein -mHIP1a
US-09-085-199B-11

  Query Match              5.2%;  Score 125.5;  DB 4;  Length 1068;
  Best Local Similarity   20.9%;  Pred. No. 0.0087;
  Matches  117;  Conservative   72;  Mismatches  206;  Indels  165;  Gaps   24;

Qy      19 RTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQI 78
           ||     |  ||  : :::::::    |::  :          |:  ||  :   :
Db     537 RTEQSGSELSSRLDTLNAEKE----ALSGVVRQR----------EAELLAAQSLVREKE 581

Qy      79 TALIETRASRLHFEGETPATIADTFAKAEKL------DRLATTTSGALRATPFAMASLLQ 132
           || : :      :||     :|:  :: : |       ::||   | | |
Db     582 EALSQEQQRSSQEKGELRGQLAEKESQEQGLRQKLLDEQLAVLRSAAAEAEAI------- 634

Qy     133 YMQPAINKGDWLPAPLKPL------TP--LISGALSGAMDQVGTKMMDRATGDLHYLSAS 184
            :| |::| |        ||      :|  |:| |   |:| |          |   ||:: |
Db     635 -LQDAVSKLD------DPLHLRCTSSPDYLVSRA-QAALDSVS----GLEQGHTQYLASS 682

Qy     185 PDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAV---------------------- 223
           |     |: |:: | |  ||    :  | |     :  :
Db     683 EDA--SALVAALTRFS-HLAADTIVNGAATSHLAPTDPADRLMDTCRECGARALELVGQL 739

Qy     224 --RTVLAPALAS--RPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGL 279
           :|||  |  |  |  :|| : ||  :          :  |:  |  : |      :
Db     740 QDQTVLRRAQPSLMRAPLQGILQLGQDL-----------KPKSLDVRQEELG-----AM 782

Qy     280 KDKEPKAQLSEENDWLEAYKAIKSAS---YSGAALNAGKRMAGLPLDMATDAMGAVRSLV 336
           |||  |  :   |  :   : : | :     ||  |    :|:    |:  || | |:| ||
Db     783 VDKEMAATSAAIEDAVRRIEDMMSQARHESSGVKLEVNERI----LNSCTDLMKAIRLLV 838

Qy     337 SASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAAL 396
           |:  |  :  :|    |    ||   ||          |: |    ||:   ||    |
Db     839 MTSTSLQKEIVESG--RGAATQQEFYAKN----------SRWTEGLISASKAVGWGATQL 886

Qy     397 --TTDPAVKKAESFIQDTVKS--TASSTTGYVADQTVK------------LAKTVKDMG 439
             : |  |     : :  | |    |:||   ||   ||              ::|| :
Db     887 VESADKVVLHMGKYEELIVCSHEIAASTAQLVAASKVKANKNSPHLSRLQECSRTVNERA 946

Qy     440 GEAITHT---------------GASLRNTVNNLRQRPAREADIEE-------------- 469
           :  |                  | ||       :   |  ::|:
Db     947 ANVVASTKSGQEQIEDRDTMDFSGLSLIKLKKQEMETQVRVLELEKTLEAERVRLGELRK 1006

Qy     470 ------GGTAASPSEIPFRP 483
                ||      | | ||
Db    1007 QHYVLAGGMGTPSEEEPSRP 1026


RESULT 14
US-09-562-737-41
; Sequence 41, Application US/09562737
; Patent No. 6428967
; GENERAL INFORMATION:
;  APPLICANT:  Herz, Joachim
;  APPLICANT:  Gotthardt, Michael
;  TITLE OF INVENTION:  LDL Receptor Signaling Pathways
;  FILE REFERENCE:  UTSW0708
;  CURRENT APPLICATION NUMBER: US/09/562,737
;  CURRENT FILING DATE:  2000-05-01
;  NUMBER OF SEQ ID NOS:  132
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 41
;   LENGTH: 1024
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: Sequence
US-09-562-737-41

  Query Match              5.1%;  Score 123;  DB 4;  Length 1024;
  Best Local Similarity   22.0%;  Pred. No. 0.014;
  Matches  108;  Conservative   64;  Mismatches  165;  Indels  154;  Gaps   23;

Qy      64 SADGQAAVDVHNAQITALIETRASRLHF--EGETPATIADTFAKAEKLDRLATTTSGALR 121
           : :|:    :     |  ::: : |: ||  ||:  :|: :     :|
Db     382 TTEGEQIAQLIAGYIDIILKKKKSKDHFGLEGDEESTMLEDSVSPKK------------ 428

Qy     122 ATPFAMASLLQYMQPAINKGDW----LPAPLKPLTPLISGALSGAMDQVGT---KMMDRA 174
               :::||     : | :     ||| ::      |||       |||:
Db     429 ------STVLQQQYNRVGKVEHGSVALPAIMR------SGASGPENFQVGSMPPAQQQIT 476

Qy     175 TGDLHYLSASPDRLHDAMAASVKRHSPSL--ARQVLDTGVAVQTYSARNAVRTVLAPALA 232
           :| :|                  : | | |  |:| | ||    :  |  | :  |
Db     477 SGQMH----------------RGHMPPLTSAQQAL-TGTINSSMQAVQAAQATLDDFDT 518

Qy     233 SRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQR------GGALVLGLKDKEPKA 286
           |  | |           |: |   |::   :   | :       | | |: |   :|
Db     519 LPPLGQDA----------ASKAWRKNKMDESKHEIHSQVDAITAGTASVVNLTAGDP-- 565

Qy     287 QLSEENDWLEAYKAIKSASYSGAALNAGKRM----------AGLPLDMATDAM-GAV--- 332
           | |:      |: : | :   :: | ::           :| || |    : |||
Db     566 ---AETDYTAVGCAVTTISSNLTEMSRGVKLLAALLEDEGGSGRPLLQAAKGLAGAVSEL 622

Qy     333 -RSLVSASSLTQNGLALAGGFAG--VGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFA 389
            ||   ||:  :  |  | |  |   |:| :   :: |||  : |: ||     |||
Db     623 LRSAQPASAEPRQNLLQAAGNVGQASGELLQQIGESDTDPHFQDALMQLAKAVASAA--- 679

Qy     390 GWTTAALTTDPAVKKAESFIQDTVKS-----------TASSTTGYVA------------ 425
           |||      | ||:|  | | |               | ||:  ||
Db     680 ----AAL-----VLKAKSVAQRTEDSGLQTQVIAAATQCALSTSQLVACTKVVAPTISSP 730

Qy     426 ---DQTVK----LAKTVKD--MGGEAITHTGASLR----------NTVNNLRQRPAREAD 466
              :| |:     :|| |:       :| |  |  ||           :| | |
Db     731 VCQEQLVEAGRLVAKAVEGCVSASQAATEDGQLLRGVGAAATAVTQALNELLQH------ 784

Qy     467 IEEGGTAASPS 477
           ::    | | |:
Db     785 VKAHATGAGPA 795


RESULT 15
US-08-809-326A-1
; Sequence 1, Application US/08809326A
; Patent No. 6165478
;  GENERAL INFORMATION:
;    APPLICANT:  Izutsu, Hiroshi
;    APPLICANT:  Obara, Kazuhiko
;    APPLICANT:  Matsumoto, Akira
;    TITLE OF INVENTION:  CHLAMYDIA PNEUMONIAE ANTIGENIC POLYPEPTIDES,
;    TITLE OF INVENTION:  FUSED PROTEINS CONTAINING THE POLYPEPTIDES, DNAS CODING
;    TITLE OF INVENTION:  THEREFOR, RECOMBINANT VECTORS CARRYING THE DNAS, TRANSFORMANT
;    TITLE OF INVENTION:  CONTAINING THE RECOMBINANT VECTORS, METHOD FOR PRODUCTION OF
;    TITLE OF INVENTION:  ANTIBODY, METHOD AND REAGENTS FOR DETECTION AND/OR MEASUREMEN
;    TITLE OF INVENTION:  OF ANTIBODY, METHOD AND AGENTS FOR DIAGNOSIS OF CHLAMYDIA
;    TITLE OF INVENTION:  PNEUMONIAE INFECTIONS, PROBES AND PRIMERS FOR DETECTION AND/O
;    TITLE OF INVENTION:  MEASUREMENT OF CHLAMYDIA PNEUMONIAE GENE, AND METHOD AND
;    TITLE OF INVENTION:  REAGENTS FOR DETECTION AND/OR MEASUREMENT OF CHLAMYDIA
;    TITLE OF INVENTION:  PNEUMONIAE GENE
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/809,326A
;      FILING DATE:  19-MAR-1997
```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 224711/94
; FILING DATE: 20-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 106006/95
; FILING DATE: 28-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 106008/95
; FILING DATE: 28-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 106009/95
; FILING DATE: 28-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 106010/95
; FILING DATE: 28-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 106011/95
; FILING DATE: 28-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Miller, Charles E.
; REGISTRATION NUMBER: 24,576
; REFERENCE/DOCKET NUMBER: 7426-043-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 488 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-809-326A-1

Query Match 5.1%; Score 122; DB 4; Length 488;
Best Local Similarity 21.2%; Pred. No. 0.0053;
Matches 102; Conservative 62; Mismatches 168; Indels 148; Gaps 22;

```
Qy      25 SLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLPPADSADGQAAVDVHNAQITALIET 84
           |::|||       | |:  | ::  ||         :|  |   |      :          :|
Db       2 SISSSS----GPDNQK--NIMSQVLTS---TPQGVPQQDKLSGNETKQIQ--------QT 44

Qy      85 RASRLHFEGETPATIADTFAKAEKLDRLATTT---------------------------- 116
           |  : : | |: ||||    |    |: ::||
Db      45 RQGK-NTEMESDATIAGASGK----DKTSSTTKTETAPQQGVAAGKESSESQKAGADTGV 99

Qy     117 SGALRATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATG 176
           |||   |    |: :  || :| :                        :| |:
Db     100 SGAAATTASNTATKIA-MQTSIEE-------------------------ASKSMESTLE 132

Qy     177 DLHYLSASPDRLHDAM---------AASVKRHSPSLAR-------QVLDTGVAVQTYSAR 220
           |  |||:  :  :|:          : | |  :| | :       :|:: |:|:
Db     133 SLQSLSAAQMKEVEAVVVAALSGKSSGSAKLETPELPKPGVTPRSEVIEIGLAL-----A 187

Qy     221 NAVRTV---LAPALASRPAVQGAVDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVL 277
           |::|:      ||::  : |   |                |   : : :     :
Db     188 KAIQTLGEATKSALSNYASTQAQAD--------QTNKLGLEKQAIKIDKEREEYQEMKAA 239

Qy     278 GLKDKEPKAQLSEENDWLEAYK-AIKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLV 336
           | |: :  :   |  : |   ||   |   |     |  :|||       | ||
Db     240 EQKSKDLEGTMDTVNTVMIAVSVAITVISIVAAIFTCGAGLAGL-------AAGAA---- 288

Qy     337 SASSLTQNGLALAGGFAGVGKLQEMATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAAL 396
               | | ||| ||       :||: ||   |   || |     |    |:     |||:
Db     289 -------VGAAAAGGAAGAAAATTVATQ-ITVQAVVQAVKQAVITAVRQAI-----TAAI 335

Qy     397 TTDPAVKKA-ESFIQDTVKSTASST-----------TGYVADQTVKLAKTVKDMGGEAIT 444
           |||   ::||:  ||: | :              |  :|    ||:| :  :  : :|
Db     336 KA--AVKSGIKAFIKTLVKAIAKAISKGISKVFAKGTQMIAKNFPKLSKVISSLTSKWVT 393
```

Search completed: January 2, 2003, 15:21:41

Job time : 20.4861 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:13:55 ; Search time 34.5355 Seconds
(without alignments)
1879.027 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINQSAQQPPGVAMESFRT.....EEGGISAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2407	100.0	487	22	AAE12603	Pseudomonas syring
2	1869.5	77.7	486	22	AAE12573	Pseudomonas syring
3	1869.5	77.7	486	22	AAB67677	Amino acid sequenc
4	147	6.1	1463	23	AAE20110	Lactobacillus rham
5	146.5	6.1	2123	22	AAE00701	Moraxella catarrha
6	140	5.8	2285	20	AAW98149	Bacillus subtilis
7	138	5.7	1229	22	ABB71736	Drosophila melanog
8	138	5.7	1277	22	AAB76536	Corynebacterium gl
9	138	5.7	2012	22	AAB76534	Corynebacterium gl
10	138	5.7	2993	22	AAG92485	C glutamicum prote
11	136	5.7	1026	15	AAR48993	rsaA S-lyaer prote
12	132	5.5	1026	18	AAW37490	Caulobacter cresce
13	132	5.5	1026	21	AAY44757	Caulobacter cresce
14	130.5	5.4	2368	22	AAU34139	Staphylococcus aur
15	130.5	5.4	2368	22	AAU36796	Staphylococcus aur
16	130	5.4	1026	17	AAR94014	Caulobacter S-laye
17	129.5	5.4	2076	22	AAU34319	Staphylococcus aur
18	129.5	5.4	2186	22	AAU37320	Staphylococcus aur
19	129	5.4	2344	22	AAU37120	Staphylococcus aur
20	129	5.4	6025	22	AAG84939	Shrimp white spot
21	127.5	5.3	2387	22	AAU01183	Rat glutamate tran
22	127.5	5.3	2478	22	AAU34320	Staphylococcus aur
23	127.5	5.3	2478	22	AAU37374	Staphylococcus aur
24	126	5.2	727	22	ABG25530	Novel human diagno
25	125.5	5.2	623	23	AAU12038	Clostridium diffic
26	125.5	5.2	1252	16	AAR80530	B. sphaericus SLP.
27	125	5.2	456	22	ABB64582	Drosophila melanog
28	125	5.2	793	22	AAU27571	Neisseria meningit
29	125	5.2	793	22	AAE10016	N. meningitidis st
30	125	5.2	2086	22	AAU34143	Staphylococcus aur
31	125	5.2	5795	22	AAU37017	Staphylococcus aur
32	124	5.2	628	23	AAG66009	F. necrophorum tru
33	124	5.2	3241	23	AAG66005	F. necrophorum leu
34	124	5.2	3647	11	AAR05041	Filamentous haemag
35	123.5	5.1	2434	22	AAU34339	Staphylococcus aur
36	123.5	5.1	6281	22	AAU37403	Staphylococcus aur
37	123	5.1	230	22	ABB58344	Drosophila melanog
38	123	5.1	232	22	ABB58317	Drosophila melanog
39	123	5.1	688	21	AAY57274	E. canis 120 kDa i
40	123	5.1	1328	20	AAY20088	B. burgdorferi ant
41	122.5	5.1	1040	23	ABP40520	Staphylococcus epi
42	121.5	5.0	1068	21	AAY59273	Mouse huntingtin-i
43	121.5	5.0	1787	23	ABB49791	Listeria monocytog
44	120.5	5.0	548	22	ABB62090	Drosophila melanog
45	120.5	5.0	810	22	AAG81232	Mycobacterium tube

ALIGNMENTS

RESULT 1
AAE12603
ID AAE12603 standard; Protein; 487 AA.
XX
AC AAE12603;
XX
DT 03-JAN-2002 (first entry)
XX
DE Pseudomonas syringae pv. tomato strain DC3000 HopPtoA2 protein.
XX
KW Conserved Effector Loci; CEL; cytostatic; antibacterial; gene therapy;
KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
KW eukaryotic cell death; cancer.
XX
OS Pseudomonas syringae.
XX
PN WO200175066-A2.
XX
PD 11-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA (UYNE-) UNIV NEBRASKA.
XX
PI Collmer A, Alfano JR, Charkowski AO;
XX
DR WPI; 2001-639361/73.

DR N-PSDB; AAD20438.
XX
PT New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
PT genomic sequences, for imparting disease resistance to plants -
XX
PS Claim 8; Page 65-66; 217pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant
CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC Pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targetted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targetted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is
CC Pseudomonas syringae pv. syringae HopPsyA homolog protein.
XX
SQ Sequence 487 AA;

Query Match 100.0%; Score 2407; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 9.4e-192;
Matches 487; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60

Qy      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120

Qy     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180

Qy     181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240

Qy     241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300

Qy     301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360

Qy     361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420

Qy     421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480

Qy     481 PFQLRRL 487
          |||||||
```

Db 481 PFQLRRL 487

RESULT 2
AAE12573
ID AAE12573 standard; Protein; 486 AA.
XX
AC AAE12573;
XX
DT 03-JAN-2002 (first entry)
XX
DE Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 encoded protein.
XX
KW Conserved Effector Loci; CEL; cytostatic; antibacterial; gene therapy;
KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
KW eukaryotic cell death; cancer.
XX
OS Pseudomonas syringae.
XX
PN WO200175066-A2.
XX
PD 11-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA (UYNE-) UNIV NEBRASKA.
XX
PI Collmer A, Alfano JR, Charkowski AO;
XX
DR WPI; 2001-639361/73.
DR N-PSDB; AAD20408.
XX
PT New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
PT genomic sequences, for imparting disease resistance to plants -
XX
PS Claim 8; Page 21-23; 217pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant
CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC Pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targetted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targetted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is
CC Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 protein.
XX
SQ Sequence 486 AA;

Query Match 77.7%; Score 1869.5; DB 22; Length 486;
Best Local Similarity 78.4%; Pred. No. 5e-147;

```
  Matches  381;  Conservative   33;  Mismatches   71;  Indels    1;  Gaps    1;

Qy       1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
           ||||:  ||||  | :||||||||||||||||||:   |:: || |||  ||||||:
Db       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60

Qy      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
             | | : |   || || ||:||||:||: |||||||||:|||||| |||||| | |
Db      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120

Qy     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180
             |||| ||:|||||||||||||  |||||||||||||||||||||||||||||| ||||
Db     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180

Qy     181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240
           || |||:|||||| |||||||:| |||:| |:||||:|| | |||||||||||||:||||
Db     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240

Qy     241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300
           || ||| |||| |||||||:|:|||||||  |||| |||:|||||||| |||| |||:||||
Db     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300

Qy     301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360
           ||||||||||||||||||||||:||| : ||||||||:|||:||||||||:||| |||:
Db     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360

Qy     361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
           |||||||| |||||||||:|| ||  |||||||| | |||||||||||||| ||||||||
Db     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420

Qy     421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480
           | |||||||||||||||| |||||: |||||||:|||||| | | |||||||| :| | ||
Db     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAA-SPSEI 479

Qy     481 PFQLRR 486
             ||:  |
Db     480 PFRPMR 485


RESULT 3
AAB67677
ID   AAB67677 standard; Protein; 486 AA.
XX
AC   AAB67677;
XX
DT   11-JUN-2001  (first entry)
XX
DE   Amino acid sequence of a HopPtoA protein of Pseudomonas syringae.
XX
KW   HIV; tat protein; effector protein; transduction domain; HopPtoA protein.
XX
OS   Pseudomonas syringae.
XX
PN   WO200119393-A1.
XX
PD   22-MAR-2001.
XX
PF   13-SEP-2000; 2000WO-US24977.
XX
PR   13-SEP-1999;   99US-0153507.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Collmer A,  Beer SV;
XX
DR   WPI; 2001-257850/26.
DR   N-PSDB; AAF55683.
XX
PT   Delivering effector proteins into target cell for use in protein
PT   therapy, involves introducing effector protein fused to protein
PT   transduction domain of human immunodeficiency virus TAT protein, into
PT   target cells -
XX
PS   Example 2; Page 23; 43pp; English.
XX
CC   The present sequence represents a HopPtoA protein of Pseudomonas syringae
CC   pv. tomato DC300 CEL.. This is an effector protein, which is used in
CC   the method of the invention. The specification describes a method for
CC   delivering effector proteins into a target cell. The method comprises
CC   introducing an effector protein fused to a protein transduction domain
CC   of a human immunodeficiency virus (HIV) tat protein into the target
CC   cell. The method is used for delivering a heterologous effector protein
CC   such as an effector protein produced by bacterial plant pathogen, animal
CC   pathogen or a rhizosphere bacteria, or a protein secreted and/or
CC   delivered into eukaryotic cells by a Type III secretion system or a
CC   hypersensitive response elicitor, an avirulence protein, a
CC   hypersensitive response and pathogenicity-dependent outer protein, a
CC   virulence protein or a pathogenicity protein, into an eukaryotic cell.
CC   The method is useful for delivering effector proteins for use in
CC   pharmaceutical, insecticidal, fungicide, herbicide and other
CC   applications.
XX
SQ   Sequence   486 AA;

  Query Match             77.7%;  Score 1869.5;  DB 22;  Length 486;
  Best Local Similarity   78.4%;  Pred. No. 5e-147;
  Matches  381;  Conservative   33;  Mismatches   71;  Indels    1;  Gaps    1;

Qy       1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
           ||||:  ||||  | :||||||||||||||||||:   |:: || |||  ||||||:
Db       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60

Qy      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
             | | : |   || || ||:||||:||: |||||||||:|||||| |||||| | |
Db      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120

Qy     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180
             |||| ||:|||||||||||||  |||||||||||||||||||||||||||||| ||||
Db     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180

Qy     181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240
           || |||:|||||| |||||||:| |||:| |:||||:|| | |||||||||||||:||||
Db     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240

Qy     241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300
           || ||| |||| |||||||:|:|||||||  |||| |||:|||||||| |||| |||:||||
Db     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300

Qy     301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360
           ||||||||||||||||||||||:||| : ||||||||:|||:||||||||:||| |||:
Db     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360

Qy     361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
           |||||||| |||||||||:|| ||  |||||||| | |||||||||||||| ||||||||
Db     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420

Qy     421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480
           | |||||||||||||||| |||||: |||||||:|||||| | | |||||||| :| | ||
Db     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAA-SPSEI 479

Qy     481 PFQLRR 486
             ||:  |
Db     480 PFRPMR 485


RESULT 4
AAE20110
ID   AAE20110 standard; Protein; 1463 AA.
XX
AC   AAE20110;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Lactobacillus rhamnosus outer membrane protein rompA.
XX
KW   Enzyme; flavour; aroma; texture; nutritional; dairy manufacture; therapy;
```

KW fermentation process; anti-infection; rotavirus infection; heart disease;
KW infantile diarrhoea; lactose digestion; anti-cancer; autoimmune disorder;
KW anti-mutagenesis; immune system modulation; allergy; Helicobacter pylori;
KW antihypertensive effect; urogenical infection; hepatic encephalopathy;
KW bowel syndrome; endocarditis; transgenic microbe; outer membrane protein;
KW rompA.
XX
OS Lactobacillus rhamnosus HN001.
XX
PN WO200212506-A1.
XX
PD 14-FEB-2002.
XX
PF 08-AUG-2001; 2001WO-NZ00160.
XX
PR 08-AUG-2000; 2000US-0634238.
PR 28-NOV-2000; 2000US-0724623.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (VIAL-) VIALACTIA BIOSCIENCE NZ LTD.
XX
PI Glenn M, Havukkala IJ, Bloksberg LN, Lubbers MW, Dekker J;
PI Christensson AC, Holland R, O'toole PW, Reid JR, Coolbear T;
XX
DR WPI; 2002-241760/29.
DR N-PSDB; AAD31881.
XX
PT New polynucleotides and polypeptides from Lactobacillus rhamnosus,
PT useful in e.g. improving the flavor, aroma, texture and health-related
PT benefits of milk-derived products, or in increasing properties of
PT microbes -
XX
PS Claim 2; Fig 64; 257pp; English.
XX
CC The present invention relates to a new isolated polynucleotide comprising
CC a sequence present in Lactobacillus rhamnosus strain HN001 and encoding a
CC polypeptide capable of modifying the flavour, aroma, texture, nutritional
CC and health benefits of milk-derived products, and/or survivability of
CC microbes in dairy manufacturing processes. The polynucleotides are useful
CC for improving the properties of microbes used in the manufacture of milk-
CC derived products such as cheeses, yogurt, fermented milk products, sour
CC milks and buttermilk; in modifying the flavour, aroma, texture and health
CC -related benefits of milk-derived products and in increasing the survival
CC of microbes during industrial fermentation processes. The bacteria may be
CC used to increase resistance to enteric pathogens and anti-infection
CC activity, including treatment of rotavirus infection and infantile
CC diarrhoea; aid in lactose digestion; as anti-cancer and anti-mutagenesis;
CC liver cancer reduction; reduction of small bowel bacterial overgrowth;
CC immune system modulation and treatment of autoimmune disorders and
CC allergies; treatment of allergic responses to foods; reduction of blood
CC lipids and prevention of heart disease; antihypertensive effect;
CC prevention and treatment of urogenical infections, Helicobacter pylori,
CC or hepatic encephalopathy; treatment of inflammatory bowel disorder and
CC irritable bowel syndrome; modulation of endocarditis; and for improved
CC protein and carbohydrate utilization and conversion. The transgenic
CC microbial population can be administered to a mammal as an anti-
CC carcinogenic agent. The present sequence is Lactobacillus rhamnosus
CC outer membrane protein rompA.
XX
SQ Sequence 1463 AA;

Query Match 6.1%; Score 147; DB 23; Length 1463;
Best Local Similarity 20.8%; Pred. No. 0.005;
Matches 104; Conservative 64; Mismatches 241; Indels 92; Gaps 12;

```
Qy      1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
Db    714 INANSAADVTASQAKATSAAASDAASYASEAQSIAGSHADNME-----IKSLASDAEKQS 768

Qy     61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATI-----AETFAKAEKFDRLATT 115
Db    769 QIALAASKSAAASSSAAASAAIV--ASSAASEASSAAAAVSNADASANSAAAAYDSYASE 826

Qy    116 ASSA--------FENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVG 167
Db    827 ASAASAANDSSGYATASFAASSAAAAM----------------------SAALSTAQVAAK 865

Qy    168 TKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVR--T 225
Db    866 VAVSDAAAAG-----------SAAAVASAAQSDSKNKQATAATARSQALDDLNKIKSLT 913

Qy    226 VLAPALASRPSVQG----------AVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGA 274
Db    914 DYASGASSSASEAGQASTATSAYASAASSSASEAGSYAHQAGSSASDAVGQSGSAAQHAS 973

Qy    275 FVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRS 334
Db    974 TAASAASSYPKDSGIQSLASQAASEAAKASSNASAATSA----AAVGFSAASDASEQAKT 1029

Qy    335 LVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTA 394
Db   1030 AASADVVASSAASTANSNASAAA----------SATKAGDSKAA---------AGFSSA 1069

Qy    395 GLATDPAVKKAESFIQDKVKSTAS--STTSYVADQTVKLAKTVKDMSGEAISSTGASLRS 452
Db   1070 ASAAASSAKGAEAVASEAASAAASDDSVASSAASAAAGFDKAASAAEGAASSAASAAASS 1129

Qy    453 TVNNLRHRSAPEADIEEGGIS 473
Db   1130 AAAQGTRGGASSSASEAGQAS 1150
```

RESULT 5
AAE00701
ID AAE00701 standard; Protein; 2123 AA.
XX
AC AAE00701;
XX
DT 02-JUL-2001 (first entry)
XX
DE Moraxella catarrhalis outer membrane protein-106 (OMP106).
XX
KW Outer membrane protein-106; OMP106; haemagglutination; vaccine;
KW bacterial infection; immunogen; cytotoxic; antibiotic;
KW passive immunisation.
XX
OS Moraxella catarrhalis.
XX
PN US6214981-B1.
XX
PD 10-APR-2001.
XX
PF 12-NOV-1997; 97US-0968685.
XX
PR 03-MAY-1996; 96US-0642712.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tucker K, Plosila L, Tillman UF;
XX
DR WPI; 2001-281002/29.
DR N-PSDB; AAD04029.
XX
PT Novel nucleotide sequences encoding Moraxella catarrhalis outer
PT membrane protein-106 polypeptide, useful for diagnosis of bacterial
PT infections and as vaccine against Moraxella catarrhalis infection of
PT mammals -
XX
PS Claim 7; Column 53-64; 49pp; English.
XX
CC The present sequence is haemagglutinating Moraxella catarrhalis outer
CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic
CC and prophylactic vaccine against M. catarrhalis infections of mammals.
CC It is used for diagnosis of bacterial infections and as reagents for
CC clinical or medical diagnosis of M. catarrhalis infections and for
CC scientific research on the properties of pathogenicity, virulence and

```
CC   infectivity of M. catarrhalis. It is also used as a probe to identify
CC   the presence of M. catarrhalis in biological specimens and to identify
CC   other bacteria that encode a polypeptide related to M. catarrhalis
CC   OMP106. OMP106-derived polypeptides are used as ligands to detect
CC   antibodies elicited in response to M. catarrhalis infections and also
CC   as immunogens for inducing M. catarrhalis-specific antibodies which are
CC   useful in immunoassays to detect M. catarrhalis in biological specimens.
CC   Cytotoxic antibodies are useful in passive immunisations against
CC   M. catarrhalis.
XX
SQ   Sequence   2123 AA;

  Query Match             6.1%;  Score 146.5;  DB 22;  Length 2123;
  Best Local Similarity   22.2%;  Pred. No. 0.0092;
  Matches  111;  Conservative   69;  Mismatches  181;  Indels  139;  Gaps   23;

Qy      32 RSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDALVETR---ANR 88
Db     377 KTLTITGGAQTSALTD---------HNIGVVQNGDGLKVQLAETLTSLKMVTTENLTANE 427

Qy      89 LYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPL 148
Db     428 KVTVGKT--------------RL-TTDKIGFTND----MNGIDESKPYLDKDTGIHAGG 467

Qy     149 KPLTPLISGAL-----------------SGAMDQVGTKMMDRARGDLHYLSTSPDKLHDA 191
Db     468 QKITKLTAGVVDDDAATYGQLKKVNQTAESALQTFTVKKVDKNGND----------ANDS 517

Qy     192 MAVSVKRHSPALGRQVVDM------GIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGV 245
Db     518 KIITVGKNNKPDGTQVNTLKLKGENGVDVTT-----------------ETNGTVTFGL 558

Qy     246 STAGGL-VANAGFGDRMLSVQ---SRDQLRGGAFVLGMKD---KEPKAALSEET----DW 294
Db     559 NQNNGLTVGNSTLNNDGLSVKNTNSNKQIQVGADGITFTDISNSKPGAGIENTTRITRDG 618

Qy     295 LDAYKAIKSASYSGAALNAGKRMAGLPLDVATD---GLKAVRSLVSATSLTKNGLALAGG 351
Db     619 I------------GFANNTGSLDANKPRLTPTGINAGGKELTNVQSAINPATNG------ 660

Qy     352 YAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAV-KKAESFIQ 410
Db     661 -GQLDFMNRLSTANTEKSGSAATIKDLYNLSQVPLTFAG------DTGPNVTKKLGEILK 713

Qy     411 DK-VKSTASSTT----SYVADQ-----TVKLAKTVKDMSGEAISSTGASLRSTVNNLRHR 460
Db     714 VKGGKTTADDLTKNNIGVVADSTDNSLTVKLAKTLSDLDAVNTKTLTASDKVTVDSGNN- 772

Qy     461 SAPEADIEEGGISAFSRSET 480
Db     773 ---TAKLQNGDLT-FSKQNT 788


RESULT 6
AAW98149
ID   AAW98149 standard; Protein; 2285 AA.
XX
AC   AAW98149;
XX
DT   05-JUL-1999  (first entry)
XX
DE   Bacillus subtilis metalloprotease YOMI.
XX
KW   YOMI; metalloprotease; protease; textile; animal feed; detergent;
KW   Gram-positive bacterium.
XX
OS   Bacillus subtilis.
XX
PN   WO9914342-A1.
XX
PD   25-MAR-1999.
XX
PF   08-SEP-1998;   98WO-US18828.
XX
PR   15-SEP-1997;   97GB-0019636.
XX
PA   (GEMV ) GENENCOR INT BV.
PA   (GEMV ) GENENCOR INT INC.
XX
PI   Estell DA;
XX
DR   WPI; 1999-229541/19.
DR   N-PSDB; AAX24980.
XX
PT   Compositions containing metalloprotease from Gram-positive
PT   microorganism
XX
PS   Claim 4; Fig 1A-O; 59pp; English.
XX
CC   The present sequence is metalloprotease YOMI of Bacillus subtilis.
CC   The invention relates to the discovery of this previously unknown
CC   metalloprotease, uses of the metalloprotease in industrial and
CC   agricultural applications, and advantageous strain improvements
CC   based on genetically engineering a Gram-positive microorganism to
CC   delete, underexpress or overexpress the enzyme.  Due to overall
CC   relatedness of the enzyme with Pseudomonas lasA protein, YOMI
CC   appears to be a member of the M23 metalloprotease family.  The
CC   metalloprotease can be used in claimed cleaning compositions,
CC   animal feed and compositions for the treatment of textiles.  It may
CC   also be used for peptide hydrolysis, waste treatment and for
CC   cleaving recombinant fusion proteins.  Expression vectors
CC   comprising a nucleic acid (see AAX24980) encoding the metalloprotease
CC   and host cells are claimed.  Gram-positive cells in which YOMI is
CC   inactivated (by gene mutation or deletion) are used for production
CC   of heterologous proteins, especially enzymes, hormones, growth
CC   factors and cytokines.
XX
SQ   Sequence   2285 AA;

  Query Match             5.8%;  Score 140;  DB 20;  Length 2285;
  Best Local Similarity   20.5%;  Pred. No. 0.036;
  Matches  128;  Conservative   80;  Mismatches  224;  Indels  192;  Gaps   26;

Qy      15 MESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSV-IGSPDERDAALA 73
Db     251 IELYQRQAQVNVQNLNTRYGSSMGSSNRQAVQDYLN----AVNSLNVSTGSNNIRSQIQS 306

Qy      74 HNEQIDAL---VETRANRLYSEGETPATIAETF-----------------------AKA 106
Db     307 LNMQFRELASNAQTAANQASSFG---AELTQTFKSMSTYLISGSLFYGAISGLKEMVSQA 363

Qy     107 EKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQV 166
Db     364 IEIDTLMTNIRRVMNEPDYKYNELL---QESIDLGDTLSNKITDILQMTGDFGRMGFDES 420

Qy     167 GTKMMDRARGDLHYLS-TSPDKLHDAMAVSVKRHSPALGRQV--------VDMGIAVQTF 217
Db     421 ELSTLTKTAQVLQNVSDLTPDDTVNTLTAAMLNFNIAANDSISIADKLNEVDNNYAVTTL 480

Qy     218 SALNVVRTVLAPALASRPSVQGAVDFGVSTA--------GGLVAN------AGFGDRMLS 263
Db     481 DLANSIRK--AGSTASTFGVELNDLIGYTTAIASTTRESGNIVGNSLKTIFARIGNNQSS 538

Qy     264 VQSRDQLRGGAFVLGMKDKEPKAALSE---ETDWL-DAYKAIKSASYSG----------- 308
Db     539 IKALEQIGISVKTAGGEAKSASDLISEVAGKWDTLSDAQKQNTSIGVAGIYQLSRFNAMM 598

Qy     309 ----AALNAGK---------------------------------------------- 315
Db     599 NNFSIAQNAAKTAANSTGSAWSEQQKYADSLQARVNKLQNNFTEFAIAASDAFISDGLIE 658

Qy     316 --RMAGLPLDVATDGLKAV---RSLVSATS-----LTKNGLALAGG-YAGVSKL--QKMA 362
Db     659 FTQAAGSLLNASTGVIKSVGFLPPLLAAVSTATLLLSKNTRTLASSLILGTRAMGQETLA 718

Qy     363 TKNITDSATKAAVSQ----------LSNLVGSVGVFAGWTTAGLATDPA-VKKA----- 405
```

```
          |  :    |:|||:           :| |||      ||    | :  |  |||
Db    719 TAGLEAGMTRAAVASRVLKTALRGLLVSTLVGGAFAALGWALESLISSFAEAKKAKDDFE 778

Qy    406 ----------------ESFIQ-----DKVKSTASSTTSYVADQ---TVKLAKT----VK 436
                         :  ||       ||| : | |:     :    | :||:|    ||
Db    779 QSQQTNVEAITTNKDSTDKLIQQYKELQKVKESRSLTSDEEQEYLQVTQQLAQTFPALVK 838

Qy    437 --DMSGEAISSTGASLRSTVNNLR 458
            |  | ||  |   |   : | :
Db    839 GYDSQGNAILKTNKELEKAIENTK 862


RESULT 7
ABB71736
ID   ABB71736 standard; Protein; 1229 AA.
XX
AC   ABB71736;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 42000.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL15839.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 42000; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1229 AA;

  Query Match             5.7%;  Score 138;  DB 22;  Length 1229;
  Best Local Similarity   20.9%;  Pred. No. 0.022;
  Matches  107;  Conservative   73;  Mismatches  205;  Indels  128;  Gaps   21;

Qy     11 PGVAMESFRTASDASLASSS--VRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDER 68
          ||:   :  |:   | :|:|  :  |:| |   | ::  |          :|   ||
Db    166 PGITKVATYTSPGYSYSSASPGISKVATYSSPSLSSVPYYAPPVTSKVETYS---SP--- 219

Qy     69 DAALAHNEQID--ALVETRANRLYSEGETPATI------------AETFAKAEKFD--- 110
            |  :::     : ||| :: || |:    |              | | ||   :
```

```
Db    220 --AYTYSKTTPGYSKVETYSSPGYSYGQISPGISRIATYSPSVSYSAPTIAKVSTYSAPS 277

Qy    111 -RLATTAS-SAFENTPFAAA---SVLQYMQPAINKGDWLATPLKPLTPLIS-------GA 158
           :||||:|  :   | ::|:   |: :| | :     ::: |:    | |:       |
Db    278 VKLATTSSLLSSHGTGYSASYAPSITKYSQVSDVSHQYISKPIVAAYPAITKVAASYGGT 337

Qy    159 LSGAMDQ------VGTKMMDRARGDLHYLSTSP--DKLHDAMAVSVKRHSPALGRQVVDM 210
           |||:          |:   |   :   |: |   ||  :   |    | |:  | |
Db    338 ASGALSHQYVSQPAIAKVSTYAAPTVATYSSGPAISKLSTSYGAS---GSGAVSHQYVSK 394

Qy    211 GIAVQTFSALNVVRTVLAPALAS---RPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSR 267
                 |:  | |  |||::|    |::     :   |     :  |:|
Db    395 PAVAIAAPAVAKVATYAAPAISSYSTGPAISKVASYAAPTVSTYSSGYGYGS-------- 446

Qy    268 DQLRGGAFVLGMKDKEPKAALSEETDWLDAYKAI-KSASYSGAALNAGKRMAGLPL---- 322
           |   |      :|  |:|       |  || | |:|:  |::     |  |:
Db    447 ---SGSGAVSHQYVSKPAVAIS-------AAPAIAKVATYAAPAIST---YAAAPVVTKV 493

Qy    323 ---------------------------DVATDGLKAVRSLVSATSLTKNGLALAGGYAGV 355
                                        |||    |: :  :| ::||   :  |   |
Db    494 ATGYGGSGSGYSSGAVSHQYVSKPAVAKVATYAAPAISTYSAAPAVTKIATSYGGSGHGA 553

Qy    356 SKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKS 415
           | ::   :  ||  |       :  |  :|    :  || | | |  ::    : :
Db    554 VSHQYVSKPAVAISAAPA--------IAKVATYASPAISTYATAPVVSKVATYAAPSIAT 605

Qy    416 TASSTTSYVADQTVKLAKT----VKDMSGEAIS 444
           :|:          |||       |:| : ||
Db    606 YSSAPA---------LAKVSYSQAADVSHQYIS 629


RESULT 8
AAB76536
ID   AAB76536 standard; Protein; 1277 AA.
XX
AC   AAB76536;
XX
DT   11-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum MCT protein SEQ ID NO:54.
XX
KW   Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW   membrane construction and membrane transport protein; petroleum spill;
KW   hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW   identification; microorganism; fine chemical production; transformation;
KW   genome mapping; genetic engineering.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100805-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB00926.
XX
PR   25-JUN-1999;   99US-0141031.
PR   08-JUL-1999;   99DE-1031454.
PR   08-JUL-1999;   99DE-1031478.
PR   08-JUL-1999;   99DE-1031563.
PR   09-JUL-1999;   99DE-1032122.
PR   09-JUL-1999;   99DE-1032124.
PR   09-JUL-1999;   99DE-1032125.
PR   09-JUL-1999;   99DE-1032128.
PR   09-JUL-1999;   99DE-1032180.
PR   09-JUL-1999;   99DE-1032182.
PR   09-JUL-1999;   99DE-1032190.
PR   09-JUL-1999;   99DE-1032191.
PR   09-JUL-1999;   99DE-1032209.
PR   09-JUL-1999;   99DE-1032212.
PR   09-JUL-1999;   99DE-1032227.
PR   09-JUL-1999;   99DE-1032228.
PR   09-JUL-1999;   99DE-1032229.
PR   09-JUL-1999;   99DE-1032230.
```

```
PR   14-JUL-1999;   99DE-1032927.
PR   14-JUL-1999;   99DE-1033005.
PR   14-JUL-1999;   99DE-1033006.
PR   27-AUG-1999;   99DE-1040764.
PR   27-AUG-1999;   99DE-1040765.
PR   27-AUG-1999;   99DE-1040766.
PR   27-AUG-1999;   99DE-1040830.
PR   27-AUG-1999;   99DE-1040831.
PR   27-AUG-1999;   99DE-1040832.
PR   27-AUG-1999;   99DE-1040833.
PR   31-AUG-1999;   99DE-1041378.
PR   31-AUG-1999;   99DE-1041379.
PR   31-AUG-1999;   99DE-1041395.
PR   03-SEP-1999;   99DE-1042077.
PR   03-SEP-1999;   99DE-1042078.
PR   03-SEP-1999;   99DE-1042079.
PR   03-SEP-1999;   99DE-1042088.
XX
PA   (BADI ) BASF AG.
XX
PI   Pompejus M,  Kroeger B,  Schroeder H,  Zelder O,  Haberhauer G;
XX
DR   WPI; 2001-071486/08.
DR   N-PSDB; AAF67769.
XX
PT   Corynebacterium glutamicum nucleic acids encoding membrane construction
PT   and membrane transport proteins or their portions, useful for typing or
PT   identifying C. glutamicum or related bacteria, and as markers for
PT   transformation -
XX
PS   Claim 20; Page 242-246; 1119pp; English.
XX
CC   AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC   construction and membrane transport (MCT) proteins given in AAB76510 to
CC   AAB76847. The MCT nucleic acids and proteins are useful in the
CC   identification of microorganisms which can be used to produce fine
CC   chemicals, for modulating fine chemical production in C. glutamicum or
CC   related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC   identification of C. glutamicum or related bacteria, as reference points
CC   for mapping C. glutamicum genome, and as markers for transformation.
CC   AAF68082 and AAF68082 represent sequencing primers which are used in an
CC   example from the present invention.
XX
SQ   Sequence   1277 AA;

  Query Match              5.7%;  Score 138;  DB 22;  Length 1277;
  Best Local Similarity   21.7%;  Pred. No. 0.023;
  Matches  119;  Conservative   85;  Mismatches  219;  Indels  126;  Gaps   23;

Qy       7 AQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPD 66
Db     313 AHTAPDVLVGRAWPAVFAAVKSAVIPGTDSASVVEGMLSLVHLEHHI-------VLKSDV 365

Qy      67 ERDAALAHNEQIDALVETRANRLY--------SEGETPATIAETFA-KAEKFDRLATTAS 117
Db     366 PTDGALKVSATADEVVDTDLGRLVIVRAEIADAEGNLIATLAERFAIRGRKGNAVARTNT 425

Qy     118 SAFE---NTPFAAASVLQYMQP------AINKGDWLATPLKPLTPL----ISGALSGAMD 164
Db     426 SALPTTVDTPRSARAVATVVAPESMRPFAVISGD--------RNPIHVSDVAASLAGLPG 477

Qy     165 QVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRH-----SPALGRQVVDMGI---AVQT 216
Db     478 VIVHGMWTSAIGELIAGAAFNDEQIQTPAAKVVEYTATMLAPVLPGEEIEFSVERSAVDN 537

Qy     217 FSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVA--NAGFGDRMLSVQSRDQLRGGA 274
Db     538 RPGMGEVRTVTA-------TVNGNL---VLTATAVVAAPSTFYAFPGQGIQSQG------ 581

Qy     275 FVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVR- 333
Db     582 --MGMEARRNSQAARAIWDRADAHTRNK------LGFSIVEIVENNPREVTVAGEKFFHP 633

Qy     334 -SLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWT 392
Db     634 DGVLYLTQFTQVGMATL-GVAQIAEMREAHALNQRAYFAGHSVGEYNALAAYAGVLSLES 692

Qy     393 T------AGLATDPAVKKAESFIQ---------DKVKSTASSTTSYVADQTVKLAKTVKD 437
Db     693 VLEIVYRRGLTMHRLVDRDENGLSNYALAALRPNKMGLTADNVFDYVA--------SVSE 744

Qy     438 MSGE--------------AISSTGASLRSTVNNLRHRSAPEADIEE--GGISAF---SRS 478
Db     745 ASGEFLEIVNYNLAGLQYAVAGTQAGL----------AALRADVENRAPGQRAFILIPGI 794

Qy     479 ETPFQLRRL 487
Db     795 DVPFHSSKL 803


RESULT 9
AAB76534
ID   AAB76534 standard; Protein; 2012 AA.
XX
AC   AAB76534;
XX
DT   11-APR-2001  (first entry)
XX
DE   Corynebacterium glutamicum MCT protein SEQ ID NO:50.
XX
KW   Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW   membrane construction and membrane transport protein; petroleum spill;
KW   hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW   identification; microorganism; fine chemical production; transformation;
KW   genome mapping; genetic engineering.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO200100805-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB00926.
XX
PR   25-JUN-1999;   99US-0141031.
PR   08-JUL-1999;   99DE-1031454.
PR   08-JUL-1999;   99DE-1031478.
PR   08-JUL-1999;   99DE-1031563.
PR   09-JUL-1999;   99DE-1032122.
PR   09-JUL-1999;   99DE-1032124.
PR   09-JUL-1999;   99DE-1032125.
PR   09-JUL-1999;   99DE-1032128.
PR   09-JUL-1999;   99DE-1032180.
PR   09-JUL-1999;   99DE-1032182.
PR   09-JUL-1999;   99DE-1032190.
PR   09-JUL-1999;   99DE-1032191.
PR   09-JUL-1999;   99DE-1032209.
PR   09-JUL-1999;   99DE-1032212.
PR   09-JUL-1999;   99DE-1032227.
PR   09-JUL-1999;   99DE-1032228.
PR   09-JUL-1999;   99DE-1032229.
PR   09-JUL-1999;   99DE-1032230.
PR   14-JUL-1999;   99DE-1032927.
PR   14-JUL-1999;   99DE-1033005.
PR   14-JUL-1999;   99DE-1033006.
PR   27-AUG-1999;   99DE-1040764.
PR   27-AUG-1999;   99DE-1040765.
PR   27-AUG-1999;   99DE-1040766.
PR   27-AUG-1999;   99DE-1040830.
PR   27-AUG-1999;   99DE-1040831.
PR   27-AUG-1999;   99DE-1040832.
PR   27-AUG-1999;   99DE-1040833.
PR   31-AUG-1999;   99DE-1041378.
PR   31-AUG-1999;   99DE-1041379.
PR   31-AUG-1999;   99DE-1041395.
PR   03-SEP-1999;   99DE-1042077.
```

PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-071486/08.
DR N-PSDB; AAF67767.
XX
PT Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation -
XX
PS Claim 20; Page 224-231; 1119pp; English.
XX
CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention.
XX
SQ Sequence 2012 AA;

Query Match 5.7%; Score 138; DB 22; Length 2012;
Best Local Similarity 21.7%; Pred. No. 0.044;
Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;

```
Qy       7 AQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPD 66
Db     995 AHTAPDVLVGRAWPAVFAAVKSAVIPGTDSASVVEGMLSLVHLEHHI-------VLKSDV 1047

Qy      67 ERDAALAHNEQIDALVETRANRLY--------SEGETPATIAETFA-KAEKFDRLATTAS 117
Db    1048 PTDGALKVSATADEVVDTDLGRLVIVRAEIADAEGNLIATLAERFAIRGRKGNAVARTNT 1107

Qy     118 SAFE---NTPFAAASVLQYMQP------AINKGDWLATPLKPLTPL----ISGALSGAMD 164
Db    1108 SALPTTVDTPRSARAVATVVAPESMRPFAVISGD--------RNPIHVSDVAASLAGLPG 1159

Qy     165 QVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRH-----SPALGRQVVDMGI---AVQT 216
Db    1160 VIVHGMWTSAIGELIAGAAFNDEQIQTPAAKVVEYTATMLAPVLPGEEIEFSVERSAVDN 1219

Qy     217 FSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVA--NAGFGDRMLSVQSRDQLRGGA 274
Db    1220 RPGMGEVRTVTA-------TVNGNL---VLTATAVVAAPSTFYAFPGQGIQSQG------ 1263

Qy     275 FVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVR- 333
Db    1264 --MGMEARRNSQAARAIWDRADAHTRNK------LGFSIVEIVENNPREVTVAGEKFFHP 1315

Qy     334 -SLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWT 392
Db    1316 DGVLYLTQFTQVGMATL-GVAQIAEMREAHALNQRAYFAGHSVGEYNALAAYAGVLSLES 1374

Qy     393 T------AGLATDPAVKKAESFIQ---------DKVKSTASSTTSYVADQTVKLAKTVKD 437
Db    1375 VLEIVYRRGLTMHRLVDRDENGLSNYALAALRPNKMGLTADNVFDYVA--------SVSE 1426

Qy     438 MSGE--------------AISSTGASLRSTVNNLRHRSAPEADIEE--GGISAF---SRS 478
Db    1427 ASGEFLEIVNYNLAGLQYAVAGTQAGL----------AALRADVENRAPGQRAFILIPGI 1476

Qy     479 ETPFQLRRL 487
```

Db 1477 DVPFHSSKL 1485

RESULT 10
AAG92485
ID AAG92485 standard; Protein; 2993 AA.
XX
AC AAG92485;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6239.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH67704.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 17; SEQ ID NO: 6239; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 2993 AA;

Query Match 5.7%; Score 138; DB 22; Length 2993;
Best Local Similarity 21.7%; Pred. No. 0.076;
Matches 119; Conservative 85; Mismatches 219; Indels 126; Gaps 23;

```
Qy       7 AQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPD 66
Db    1025 AHTAPDVLVGRAWPAVFAAVKSAVIPGTDSASVVEGMLSLVHLEHHI-------VLKSDV 1077

Qy      67 ERDAALAHNEQIDALVETRANRLY--------SEGETPATIAETFA-KAEKFDRLATTAS 117
Db    1078 PTDGALKVSATADEVVDTDLGRLVIVRAEIADAEGNLIATLAERFAIRGRKGNAVARTNT 1137

Qy     118 SAFE---NTPFAAASVLQYMQP------AINKGDWLATPLKPLTPL----ISGALSGAMD 164
```

```
Db   1138 SALPTTVDTPRSARAVATVVAPESMRPFAVISGD--------RNPIHVSDVAASLAGLPG 1189

Qy    165 QVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRH-----SPALGRQVVDMGI---AVQT 216
          :    |   | |:|    :    |:      |  |  :     :| |  : ::  :    ||
Db   1190 VIVHGMWTSAIGELIAGAAFNDEQIQTPAAKVVEYTATMLAPVLPGEEIEFSVERSAVDN 1249

Qy    217 FSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVA--NAGFGDRMLSVQSRDQLRGGA 274
          :  |||| |        :| | :   | ||  :||  :  :       :||:
Db   1250 RPGMGEVRTVTA-------TVNGNL---VLTATAVVAAPSTFYAFPGQGIQSQG------ 1293

Qy    275 FVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVR- 333
          :||: :    |      |  ||:   |          :  : :   | :|   | |
Db   1294 --MGMEARRNSQAARAIWDRADAHTRNK------LGFSIVEIVENNPREVTVAGEKFFHP 1345

Qy    334 -SLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWT 392
           ::  |  |: |:|   | | ::::::    |        :| : : |     || :  :
Db   1346 DGVLYLTQFTQVGMATL-GVAQIAEMREAHALNQRAYFAGHSVGEYNALAAYAGVLSLES 1404

Qy    393 T------AGLATDPAVKKAESFIQ---------DKVKSTASSTTSYVADQTVKLAKTVKD 437
                ||      | : |: :           :|:  || :   |||        :| :
Db   1405 VLEIVYRRGLTMHRLVDRDENGLSNYALAALRPNKMGLTADNVFDYVA--------SVSE 1456

Qy    438 MSGE--------------AISSTGASLRSTVNNLRHRSAPEADIEE--GGISAF---SRS 478
          |||               |:: | | |          :|  ||:|    |  ||
Db   1457 ASGEFLEIVNYNLAGLQYAVAGTQAGL----------AALRADVENRAPGQRAFILIPGI 1506

Qy    479 ETPFQLRRL 487
          : ||   :|
Db   1507 DVPFHSSKL 1515


RESULT 11
AAR48993
ID   AAR48993 standard; Protein; 1026 AA.
XX
AC   AAR48993;
XX
DT   14-SEP-1994  (first entry)
XX
DE   rsaA S-lyaer protein.
XX
KW   C. crescentus; rsaA; paracrystalline; S-layer; protein; heterologous;
KW   cellulase; xylase; metallothionein; restriction site;
KW   reading frame; fusion protein; bioreactor; toxic metal; sewage;
KW   waste water; wood pulp suspension; cell surface; vaccine; fish.
XX
OS   Caulobacter crescentus.
XX
PN   CA2090549-A.
XX
PD   10-DEC-1993.
XX
PF   26-FEB-1993;   93CA-2090549.
XX
PR   09-JUN-1992;   92US-0895367.
XX
PA   (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI   Bingle WH,  Smit J;
XX
DR   WPI; 1994-066249/09.
DR   N-PSDB; AAQ57972.
XX
PT   Prodn. of heterologous polypeptides in bacteria, partic.
PT   Caulobacter - by expression of a fusion prod. of the polypeptide
PT   sequence and a bacterial S-layer protein gene
XX
PS   Claim 17; Fig 6; 27pp; English.
XX
CC   This sequence is encoded by the C. crescentus rsaA gene and represents
CC   the paracrystalline S-layer protein. The rsaA gene was used in the
CC   production of the heterologous protein of the invention.  The
CC   heterologous protein is produced by cloning a polypeptide coding
CC   sequence, eg. cellulase, xylase or a metallothionein, into a
CC   restriction site within the rsaA gene which preserves the rsaA reading
CC   frame and expressing the fusion sequence in Caulobacter.  This S-
CC   layer protein bacterial system can be used in bioreactors, eg. to
CC   bind toxic metals in sewage waste water etc. or for the treatment of
CC   wood pulp suspensions.  The system can be used to produce heterologous
CC   proteins at the cell surface for use in vaccines, partic. fish
CC   vaccines.  The S-layer protein is synthesised in large quantities and
CC   has a general repetitive sequence, permitting the synthesis of large
CC   amounts of heterologous protein as a fusion product and presentation at
CC   the cell surface.
XX
SQ   Sequence   1026 AA;

  Query Match             5.7%;  Score 136;  DB 15;  Length 1026;
  Best Local Similarity   21.2%;  Pred. No. 0.025;
  Matches  102;  Conservative   75;  Mismatches  165;  Indels  140;  Gaps   23;

Qy      22 SDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDAL 81
           |||:  ::::: |::|:   :|       :  |      |  |             :| |
Db      42 SDAAALTNTLKLVNSTTAVAIQT------YQFFTGVAPSAAG-------------LDFL 81

Qy      82 VETRAN------RLYSEGETPATIAETFAKAEKF----DRLATTASSAFENTPFAAA--- 128
           |::  |        ||:          ||:  :|      ||| | :    | ||||
Db      82 VDSTTNTNDLNDAYYSK----------FAQENRFINFSINLATGAGAG--ATAFAAAYTG 129

Qy     129 -SVLQYMQPAINK--GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSP 185
            |  | :  | :|  |: :||              :  :|          : :: ||:
Db     130 VSYAQTVATAYDKIIGNAVAT------------AAGVDVAAAVAFLSRQANIDYLT--- 173

Qy     186 DKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVL---APALASRPSVQGAVD 242
                   |: ::|      :|: :      :    ||      | |:   :    |
Db     174 --------AFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAAMINDLSD 225

Qy     243 FGVST--AGGL-----VANAGFGDRMLSVQSR-DQLRG------------GAFVLGMKD 281
             :||  | |:       ::|     ||: :| | | |              ||  | : |
Db     226 GALSTDNAAGVNLFTAYPSSGVSGSTLSLTTRTDTLTGTANNDTFVAGEVAGAATLTVGD 285

Qy     282 KEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSL 341
                 | ::  :|: |           ||: |      ||  |    |:: : :: |  ::
Db     286 TLSGGAGTDVLNWVQA----------AAVTA------LPTGVTISGIETM-NVTSGAAI 327

Qy     342 TKNGLALAGGYAGVSKL---------------QKMATKNITDSATKAAVSQLSNL-VGS 384
           | |    : |  |:: |                 | :       :|   ||   :|: | |
Db     328 TLN---TSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGGANVTVAS 384

Qy     385 VGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAIS 444
           || :| || |     |   |   :   | :::::||  :|  |   | ||   :| |::
Db     385 TGVTSGTTTVG-----ANSAASGTVSVSVANSSTTTTGAIA-VTGGTAVTVAQTAGNAVN 438

Qy     445 ST 446
           :|
Db     439 TT 440


RESULT 12
AAW37490
ID   AAW37490 standard; Protein; 1026 AA.
XX
AC   AAW37490;
XX
DT   20-APR-1998  (first entry)
XX
DE   Caulobacter crescentus S-layer rsaA protein.
XX
KW   S-layer; rsaA gene; Caulobacter; vaccine; antigenic; ligand; enzyme;
KW   metallothionein; heavy metal; water; sewage; xylanase; cellulase;
KW   wood pulping.
XX
OS   Caulobacter crescentus.
XX
PN   WO9734000-A1.
XX
```

PD 18-SEP-1997.
XX
PF 10-MAR-1997; 97WO-CA00167.
XX
PR 12-MAR-1996; 96US-0614377.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Bingle WH, Nomellini JF, Smit J;
XX
DR WPI; 1997-470880/43.
DR N-PSDB; AAV01866.
XX
PT New DNA containing sequence for C-terminal region of Caulobacter
PT S-layer protein - expressed as fusion proteins containing antigenic
PT peptides in Caulobacter, useful as live vaccines
XX
PS Example 2; Fig 6; 58pp; English.
XX
CC The present sequence represents a Caulobacter S-layer protein used in an
CC example of the present invention. A new DNA construct has been developed
CC which contains at least one restriction site for insertion of DNA
CC upstream of DNA encoding a C-terminal region of at least the last 82
CC amino acids (aa) of Caulobacter S-layer protein. Caulobacter containing
CC the DNA constructs above additionally containing a sequence encoding a
CC heterologous polypeptide, are particularly useful in live vaccines
CC (where the heterologous polypeptide is an antigen). They can also be
CC used for production of e.g. ligands, enzymes or other proteins, e.g.
CC metallothioneins to remove heavy metals from water or sewage, or
CC xylanase or cellulase for use in wood pulping. All known Caulobacter
CC strains are harmless, and stable in outdoor environments, including
CC water (so suitable for vaccinating fish) or soil. They are well suited
CC for growing in biofilm reactors and produce S-layer proteins, which is
CC an ideal system for presentation of antigens, at high level.
XX
SQ Sequence 1026 AA;

Query Match 5.5%; Score 132; DB 18; Length 1026;
Best Local Similarity 21.0%; Pred. No. 0.054;
Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;

```
Qy      22 SDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDAL 81
           |||:  ::::: |::|:   :|       :  |     |  |                :| |
Db      42 SDAAALTNTLKLVNSTTAVAIQT------YQFFTGVAPSAAG--------------LDFL 81

Qy      82 VETRAN------RLYSEGETPATIAETFAKAEKF----DRLATTASSAFENTPFAAA--- 128
           |::  |         ||:          ||:  :|      |||  | :    | ||||
Db      82 VDSTTNTNDLNDAYYSK----------FAQENRFINFSINLATGAGAG--ATAFAAAYTG 129

Qy     129 -SVLQYMQPAINK--GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSP 185
            |  | :  | :|  |: :||           :  :|          : :: ||:
Db     130 VSYAQTVATAYDKIIGNAVAT------------AAGVDVAAAVAFLSRQANIDYLT--- 173

Qy     186 DKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVL---APALASRPSVQGAVD 242
                   |: ::|      :|: :       :    ||      | |:   :    |
Db     174 --------AFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAAMINDLSD 225

Qy     243 FGVST--AGGL-----VANAGFGDRMLSVQS-RDQLRG------------GAFVLGMKD 281
             :||  | |:        ::|      ||: :  | | |              ||  | : |
Db     226 GALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVAGEVAGAATLTVGD 285

Qy     282 KEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSL 341
                | ::  :|: |            ||: |      ||  |   |:: : :: |  ::
Db     286 TLSGGAGTDVLNWVQA----------AAVTA------LPTGVTISGIETM-NVTSGAAI 327

Qy     342 TKNGLALAGGYAGVSKL---------------QKMATKNITDSATKAAVSQLSNL-VGS 384
           | |    : |  |:: |                | :       :|   ||   :|: | |
Db     328 TLN---TSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGRANVTVAS 384

Qy     385 VGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAIS 444
            || :| || |     |   |   :  . | :::::||  :|  |   | ||   :| |::
Db     385 TGVTSGTTTVG-----ANSAASGTVSVSVANSSTTTTGAIA-VTGGTAVTVAQTAGNAVN 438
```

Qy 445 ST 446
:|
Db 439 TT 440

RESULT 13
AAY44757
ID AAY44757 standard; Protein; 1026 AA.
XX
AC AAY44757;
XX
DT 04-MAY-2000 (first entry)
XX
DE Caulobacter crescentus surface layer protein.
XX
KW Surface layer protein; S-layer secretion signal; antibiotic; vaccine;
KW recombinant fusion protein cleavage; enzyme; protein polymer;
KW antibacterial enzyme; foodstuff.
XX
OS Caulobacter crescentus.
XX
FH Key Location/Qualifiers
FT Cleavage-site 692..693
FT /note= "Asp-Pro dipeptide present in S-layer secretion
FT signal sequence. It is a site where a fusion
FT protein comprising a target protein and the secretion
FT signal is cleaved"
XX
PN WO200004170-A1.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-CA00637.
XX
PR 14-JUL-1998; 98CA-2237704.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Smit J;
XX
DR WPI; 2000-182434/16.
DR N-PSDB; AAZ50079.
XX
PT Cleavage of Caulobacter produced recombinant fusion proteins useful for
PT producing vaccine peptides -
XX
PS Example 1; Pages 21-23; 33pp; English.
XX
CC The patent discloses a method for cleaving a recombinant fusion protein
CC which is produced by Caulobacter and consists of Caulobacter surface
CC layer (S-layer) protein (containing the C-terminal secretion signal) and
CC a target protein heterologous to Caulobacter. The cleavage of target
CC protein from the S-layer protein is carried out under mild acid
CC conditions so that cleavage occurs at aspartate-
CC proline dipeptide site without solubilising the protein.
CC The cleavage is accomplished while the fusion protein is in
CC an insoluble aggregate form which facilitates purification of the
CC protein. The method is useful for producing pure proteins including
CC recombinant human and animal therapeutic antibiotic and vaccine peptides,
CC enzymes, protein polymers, and antibacterial enzymes for foodstuffs.
CC The present sequence is a S-layer protein from C. crescentus.
CC The S-layer secretion signal, corresponding to the C-terminal portion of
CC the protein from amino acid 690 onwards, is fused with a target sequence
CC for construction of a recombinant fusion construct which is expressed in
CC Caulobacter.
XX
SQ Sequence 1026 AA;

Query Match 5.5%; Score 132; DB 21; Length 1026;
Best Local Similarity 21.0%; Pred. No. 0.054;
Matches 101; Conservative 75; Mismatches 166; Indels 140; Gaps 23;

Qy 22 SDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDAL 81

```
Db     42 SDAAALTNTLKLVNSTTAVAIQT------YQFFTGVAPSAAG--------------LDFL 81

Qy     82 VETRAN------RLYSEGETPATIAETFAKAEKF----DRLATTASSAFENTPFAAA--- 128
Db     82 VDSTTNTNDLNDAYYSK----------FAQENRFINFSINLATGAGAG--ATAFAAAYTG 129

Qy    129 -SVLQYMQPAINK--GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSP 185
Db    130 VSYAQTVATAYDKIIGNAVAT-------------AAGVDVAAAVAFLSRQANIDYLT--- 173

Qy    186 DKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVL---APALASRPSVQGAVD 242
Db    174 --------AFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAAMINDLSD 225

Qy    243 FGVST--AGGL-----VANAGFGDRMLSVQS-RDQLRG-------------GAFVLGMKD 281
Db    226 GALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVAGEVAGAATLTVGD 285

Qy    282 KEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSL 341
Db    286 TLSGGAGTDVLNWVQA----------AAVTA------LPTGVTISGIETM-NVTSGAAI 327

Qy    342 TKNGLALAGGYAGVSKL---------------QKMATKNITDSATKAAVSQLSNL-VGS 384
Db    328 TLN---TSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGRANVTVAS 384

Qy    385 VGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAIS 444
Db    385 TGVTSGTTTVG-----ANSAASGTVSVSVANSSTTTTGAIA-VTGGTAVTVAQTAGNAVN 438

Qy    445 ST 446
Db    439 TT 440
```

RESULT 14
AAU34139
ID AAU34139 standard; Protein; 2368 AA.
XX
AC AAU34139;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #415.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS51998.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 5635; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2368 AA;

Query Match 5.4%; Score 130.5; DB 22; Length 2368;
Best Local Similarity 20.2%; Pred. No. 0.23;
Matches 99; Conservative 64; Mismatches 223; Indels 105; Gaps 17;

```
Qy      4 NQSAQQPPGVAMESFRTAS-------DASLASSSVRSVSTTSCRDLQAITDYLKHHVFAA 56
Db    702 NASLQDEKDVANDKIGKIETKAIKDIDAATTNAQVEAIKTKAINDINQTAPATTAKAAAL 761

Qy     57 HRFSVIGSPDERDAAL---AHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLA 113
Db    762 EEFDEVVQAQIDQAPLNPDTTNEEVAEAIE----RINAAKVSGVKAIEATTTAQDLERVK 817

Qy    114 TTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDR 173
Db    818 NEEISKIEN----ITDSTQTKMDAYNEVKQAATARKTQNATVSNATN---EEV------- 863

Qy    174 ARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALAS 233
Db    864 AEADAAVEAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAIQT---------------- 906

Qy    234 RPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETD 293
Db    907 QAKVKPAADTEVENA--------YNTRKQEIQNSN---------ASTTEEKQAAYTE--- 946

Qy    294 WLDAYKAIKSASYSGAALNAGKRMAGLPLDVAT---DGLKAVRSLVSATSLTKNGLALAG 350
Db    947 -LDTKKQEARTNLDAANTNS---------DVTTAKDNGIAAINQVQAATTKKSDAKAEIA 996

Qy    351 GYAGVSKL----------QKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTT-----A 394
Db    997 QKASERKTAIEAMNDSTTEEQQAAKDKVDQAVVTANADIDNAAANTDVDNAKTTNEATIA 1056

Qy    395 GLATDPAVK-KAESFIQDKVKST---------ASSTTSYVADQTVKLAKTVKDMSGEAIS 444
Db   1057 AITPDANVKPTAKQAIADKVQAQETAIDANNGATTEEKAAAKQQVQTEKTTADTAIDG-A 1115

Qy    445 STGASLRSTVN 455
Db   1116 HTNAEVEAAKN 1126
```

RESULT 15
AAU36796
ID AAU36796 standard; Protein; 2368 AA.
XX
AC AAU36796;

XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #966.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS54655.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 12389; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2368 AA;

Query Match 5.4%; Score 130.5; DB 22; Length 2368;
Best Local Similarity 20.2%; Pred. No. 0.23;
Matches 99; Conservative 64; Mismatches 223; Indels 105; Gaps 17;

```
Qy      4 NQSAQQPPGVAMESFRTAS-------DASLASSSVRSVSTTSCRDLQAITDYLKHHVFAA 56
          | | |    || :              ||:  :: | :: | :  |:             |
Db    702 NASLQDEKDVANDKIGKIETKAIKDIDAATTNAQVEAIKTKAINDINQTAPATTAKAAAL 761

Qy     57 HRFSVIGSPDERDAAL---AHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLA 113
            |  :       | |     ||::   :|    |: :   :     |    |:  :|:
Db    762 EEFDEVVQAQIDQAPLNPDTTNEEVAEAIE----RINAAKVSGVKAIEATTTAQDLERVK 817

Qy    114 TTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDR 173
              |  ||          |    | |:    ||  |     :| | :   ::|

Db    818 NEEISKIEN----ITDSTQTKMDAYNEVKQAATARKTQNATVSNATN---EEV------- 863

Qy    174 ARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALAS 233
          |  |    :     |||   |  |:       :|:|   |:||
Db    864 AEADAAVEAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAIQT---------------- 906

Qy    234 RPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETD 293
          :  |: | |  |  |        :  |   :|: :             :| :|| :|
Db    907 QAKVKPAADTEVENA--------YNTRKQEIQNSN---------ASTTEEKQAAYTE--- 946

Qy    294 WLDAYKAIKSASYSGAALNAGKRMAGLPLDVAT---DGLKAVRSLVSATSLTKNGLALAG 350
           ||  |     :   |  |:          || |   :|: |:  : :||:   :  |
Db    947 -LDTKKQEARTNLDAANTNS---------DVTTAKDNGIAAINQVQAATTKKSDAKAEIA 996

Qy    351 GYAGVSKL----------QKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTT-----A 394
            |   |            :: | |:  | |   | : : |   :  |    ||     |
Db    997 QKASERKTAIEAMNDSTTEEQQAAKDKVDQAVVTANADIDNAAANTDVDNAKTTNEATIA 1056

Qy    395 GLATDPAVK-KAESFIQDKVKST---------ASSTTSYVADQTVKLAKTVKDMSGEAIS 444
            :  |  ||  |:  | |||::          |::     | | |:  ||  | : :  :
Db   1057 AITPDANVKPTAKQAIADKVQAQETAIDANNGATTEEKAAAKQQVQTEKTTADTAIDG-A 1115

Qy    445 STGASLRSTVN 455
            | | : :  |
Db   1116 HTNAEVEAAKN 1126
```

Search completed: January 2, 2003, 15:19:00
Job time : 42.5355 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:16:50 ; Search time 30.0308 Seconds
(without alignments)
3341.397 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINQSAQQPPGVAMESFRT.....EEGGISAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1869.5	77.7	486	2	Q9JP34	Q9jp34 pseudomonas
2	1591.5	66.1	484	2	Q8RP03	Q8rp03 pseudomonas
3	460	19.1	518	16	Q8XQF0	Q8xqf0 ralstonia s
4	149.5	6.2	2155	16	Q8UFP9	Q8ufp9 agrobacteri
5	149	6.2	1009	16	Q8U8W1	Q8u8w1 agrobacteri
6	146	6.1	2089	16	Q92K98	Q92k98 rhizobium m
7	143.5	6.0	2055	2	O85472	O85472 abiotrophia
8	141	5.9	536	17	Q9HPA2	Q9hpa2 halobacteri
9	140.5	5.8	642	16	Q8XSW6	Q8xsw6 ralstonia s
10	140.5	5.8	1983	16	Q931N9	Q931n9 staphylococ
11	140.5	5.8	2186	16	Q99TB0	Q99tb0 staphylococ
12	140	5.8	503	2	Q9FCQ5	Q9fcq5 pseudomonas
13	140	5.8	2285	9	O64046	O64046 bacteriopha
14	140	5.8	2285	16	O31976	O31976 bacillus su
15	139.5	5.8	491	2	Q93GT1	Q93gt1 campylobact
16	139.5	5.8	1795	2	Q9LCJ9	Q9lcj9 staphylococ
17	138.5	5.8	491	2	Q8RTY4	Q8rty4 campylobact
18	138.5	5.8	548	2	Q9RA74	Q9ra74 streptococc
19	138	5.7	1229	5	Q9VW05	Q9vw05 drosophila
20	138	5.7	1731	16	Q8U8W4	Q8u8w4 agrobacteri
21	137.5	5.7	2283	2	Q8VQ99	Q8vq99 staphylococ
22	137	5.7	1545	16	Q9RDQ1	Q9rdq1 streptomyce
23	135	5.6	973	16	Q8XDQ4	Q8xdq4 escherichia
24	135	5.6	6077	12	Q8VAP1	Q8vap1 white spot
25	135	5.6	6077	12	Q8QTB7	Q8qtb7 white spot
26	134.5	5.6	1713	3	Q8TGE1	Q8tge1 saccharomyc
27	134.5	5.6	2016	5	Q9BIT0	Q9bit0 plectreurys
28	134.5	5.6	2178	2	Q9KWR3	Q9kwr3 streptococc
29	134	5.6	2535	16	Q8ZDR6	Q8zdr6 yersinia pe
30	133.5	5.5	1233	16	Q92SD7	Q92sd7 rhizobium m
31	133.5	5.5	2124	16	Q98M03	Q98m03 rhizobium l
32	132.5	5.5	654	16	Q9A4U7	Q9a4u7 caulobacter
33	132	5.5	6077	12	Q91L85	Q91l85 white spot
34	131.5	5.5	553	2	Q9L2M2	Q9l2m2 rhodobacter
35	131.5	5.5	1248	16	Q8UI01	Q8ui01 agrobacteri
36	131.5	5.5	1292	10	Q9LPN4	Q9lpn4 arabidopsis
37	131.5	5.5	1309	10	Q9CAF4	Q9caf4 arabidopsis
38	131.5	5.5	3501	16	Q8Y106	Q8y106 ralstonia s
39	131	5.4	409	16	Q56025	Q56025 salmonella
40	130	5.4	1953	16	Q98HJ2	Q98hj2 rhizobium l
41	129.5	5.4	1794	9	Q9T1A7	Q9t1a7 bacteriopha
42	129.5	5.4	3072	2	Q939N5	Q939n5 streptococc
43	129	5.4	637	16	Q9AAK8	Q9aak8 caulobacter
44	128	5.3	546	2	Q93G07	Q93g07 lactobacill
45	128	5.3	2432	12	Q8QZQ6	Q8qzq6 chilo iride

ALIGNMENTS

RESULT 1
Q9JP34
ID Q9JP34 PRELIMINARY; PRT; 486 AA.
AC Q9JP34;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 50.7 kDa protein.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=98422476; PubMed=9748456;
RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA Collmer A.;
RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
RT to harpins and pectate lyases and can elicit the plant hypersensitive
RT response and bind to pectate.";
RL J. Bacteriol. 180:5211-5217(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RX MEDLINE=20243785; PubMed=10781092;
RA Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
RA Petnicki-Ocwieja T., van Dijk K., Collmer A.;
RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT mosaic structure composed of a cluster of type III secretion genes
RT bounded by exchangeable effector and conserved effector loci that
RT contribute to parasitic fitness and pathogenicity in plants.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Ramos A.R., Rehm A.H., Collmer A.R.;
RT "Pseudomonas syringae pv. tomato DC3000 hrpL through hrcU.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.

```
DR   EMBL; AF232004; AAF71504.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   486 AA;  50709 MW;  91CFBC03397F8AD1 CRC64;

  Query Match             77.7%;  Score 1869.5;  DB 2;  Length 486;
  Best Local Similarity   78.4%;  Pred. No. 1.5e-98;
  Matches  381;  Conservative   33;  Mismatches   71;  Indels    1;  Gaps    1;

Qy       1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
Db       1 MHINRRVQQPPVTATDSFRTASDASLASSSVRSVSSDQQREINAIADYLTDHVFAAHKLP 60

Qy      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
Db      61 PADSADGQAAVDVHNAQITALIETRASRLHFEGETPATIADTFAKAEKLDRLATTTSGAL 120

Qy     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180
Db     121 RATPFAMASLLQYMQPAINKGDWLPAPLKPLTPLISGALSGAMDQVGTKMMDRATGDLHY 180

Qy     181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240
Db     181 LSASPDRLHDAMAASVKRHSPSLARQVLDTGVAVQTYSARNAVRTVLAPALASRPAVQGA 240

Qy     241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300
Db     241 VDLGVSMAGGLAANAGFGNRLLSVQSRDHQRGGALVLGLKDKEPKAQLSEENDWLEAYKA 300

Qy     301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360
Db     301 IKSASYSGAALNAGKRMAGLPLDMATDAMGAVRSLVSASSLTQNGLALAGGFAGVGKLQE 360

Qy     361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
Db     361 MATKNITDPATKAAVSQLTNLAGSAAVFAGWTTAALTTDPAVKKAESFIQDTVKSTASST 420

Qy     421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480
Db     421 TGYVADQTVKLAKTVKDMGGEAITHTGASLRNTVNNLRQRPAREADIEEGGTAA-SPSEI 479

Qy     481 PFQLRR 486
Db     480 PFRPMR 485


RESULT 2
Q8RP03
ID   Q8RP03       PRELIMINARY;      PRT;   484 AA.
AC   Q8RP03;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Type III effector HopPtoA1Pma.
OS   Pseudomonas syringae pv. maculicola.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=59511;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ES4326;
RX   MEDLINE=21862332; PubMed=11872842;
RA   Guttman D.S., Vinatzer B.A., Sarkar S.F., Ranall M.V., Kettler G.,
RA   Greenberg J.T.;
RT   "A Functional Screen for the Type III (Hrp) Secretome of the Plant
RT   Pathogen Pseudomonas syringae.";
RL   Science 295:1722-1726(2002).
DR   EMBL; AF458051; AAL84253.1; -.
SQ   SEQUENCE   484 AA;  50520 MW;  4DB012DD5688163E CRC64;

  Query Match             66.1%;  Score 1591.5;  DB 2;  Length 484;
  Best Local Similarity   67.6%;  Pred. No. 9.7e-83;
  Matches  330;  Conservative   51;  Mismatches  100;  Indels    7;  Gaps    3;

Qy       1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
Db       1 MYINRSISSQSSIGTESFHSAQ--SVASSSVRVLSEAEQTKLNDITHYLTDHVFAAHRLP 58

Qy      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
Db      59 VNDINVDQDVVHAHNVEIGKIISARALRLLDEGETALSIGETFAKAEKFDRMASAASGAL 118

Qy     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180
Db     119 RATPFAAASVLQYMQPAINKGDWLPTPLKPLAPLVSGALSGVMDQVGTGVMNRATGDLHY 178

Qy     181 LSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGA 240
Db     179 LSTAPEKLHDAMAASVKRHKPGVMRQAADLGIAVQTYTARNALRTVLAPALASRPAVQGA 238

Qy     241 VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKA 300
Db     239 VDISVSAAGSLVANAGFSERMHTVQARDHLRGGAFVLGIKDKQPKADLSEETDWLDAYRA 298

Qy     301 IKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQK 360
Db     299 IKSASYSGAALNAGKRVAGLPLDVATDGLKAVRSLVSATSLMQNGLVMAGGFAGVGKLQE 358

Qy     361 MATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASST 420
Db     359 MATKNITHPATKAAVSQLTNLAGSAPVFSAWTTAAVATDPAVKKAESFLQDTVKTTVSNG 418

Qy     421 TSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRS-- 478
Db     419 TSALADKTVKLAKAGIDASAERIATTGASLRDT---LRRRTVREPDIEEGGVAAGTGSPV 475

Qy     479 ETPFQLRR 486
Db     476 AVPFEAMR 483


RESULT 3
Q8XQF0
ID   Q8XQF0       PRELIMINARY;      PRT;   518 AA.
AC   Q8XQF0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative transmembrane protein.
GN   RSP1277 OR RS05322.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OG   Plasmid megaplasmid.
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646083; CAD18428.1; -.
KW   Plasmid; Complete proteome.
SQ   SEQUENCE   518 AA;  53581 MW;  82D486715F0F60EE CRC64;

  Query Match             19.1%;  Score 460;  DB 16;  Length 518;
  Best Local Similarity   32.0%;  Pred. No. 2.3e-18;
  Matches  139;  Conservative   63;  Mismatches  159;  Indels   74;  Gaps   12;

Qy      16 ESFRTASDASLASSSVRSVSTTSCRDLQA----------------------------ITD 47
```

```
Db      97 QSERSNSDVSQWFSAAASIRAQSIGSIESNSPPGGSPPHTADPGQTGRGIDIPHWMPIAD 156

Qy      48 YLKHHVFAAHRFSVIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAE 107

Db     157 QLR-DIFRPH---FDGTND---------AQFEILIQQRAERLQAMGETPATVAAVLAKGA 203

Qy     108 KFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLIS---GALSGAMD 164

Db     204 NRDRLAQTTVGFVRSVPFGIASRLFDVKQAL-------TAFAKTTARVGATVGAGSGTAD 256

Qy     165 QVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVR 224

Db     257 AFGGTLLGKATSNTQWLAASPDHLEPVMAQAHKAVQPSLGRLAAEVSLAFQTYSLRNVIR 316

Qy     225 TVLAP----ALASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMK 280

Db     317 TGVAPLATHALGAARAAN--VDSWIAAVGGPVAGAAAYMAMQHMNETHHRTGAEYLLG-- 372

Qy     281 DKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPL-DVATDGLKAVRSLVSAT 339

Db     373 ----------RTDWEDQFKQLKQSTWTDPLVGAGKRTAKLLVADLATETLAATRSLFTAT 422

Qy     340 SLTKNGLALAGGYAGVSKLQKMATKNITDSA-TKAAVSQLSNLVGSV---GVFAGWTTAG 395

Db     423 NLIKNMGALAGGFAGVLTAQTAAGKAATEAGYTEAAVAAVRRAVSTVLSAPVYAAWTTAD 482

Qy     396 LATDPAVKKAESFIQ 410

Db     483 VMAGPAIDAAAGHIQ 497


RESULT 4
Q8UFP9
ID   Q8UFP9      PRELIMINARY;      PRT;  2155 AA.
AC   Q8UFP9;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Atu1348.
GN   ATU1348 OR AGR_C_2490.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009096; AAL42354.1; -.
DR   EMBL; AE008061; AAK87140.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   2155 AA;  230143 MW;  FD28CCC6242ADD10 CRC64;

  Query Match             6.2%;  Score 149.5;  DB 16;  Length 2155;
  Best Local Similarity   20.7%;  Pred. No. 8;
  Matches   96;  Conservative   65;  Mismatches  176;  Indels  127;  Gaps   16;

Qy      69 DAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF----ENTP 124

Db     609 DAFTNSHAKIDTVLAERSNALF---------GALSASQDRFDEALASRSLAITGSVSGTA 659

Qy     125 FAAASVLQYMQPAINK-----GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLH 179

Db     660 EHLAAMLDERAAAINSVVADVERRLTETLETRAAAITGAVSGIEDRISDTLESRTAA--- 716

Qy     180 YLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASR----- 234

Db     717 --------LHDVVSGAESRIADTLD------GRTAALSSAISGVEERIADTMDSRTLSLD 762

Qy     235 -------PSVQGAVDFGVSTAGGLVAN-------------AGFGDRMLSVQSR--DQLRG 272

Db     763 MTFANVEERLSETLDNRTSALTGIVASAEEKIAGALDSRTATFGDVVAGAETRIAETLDG 822

Qy     273 -----GAFVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGA------------------ 309

Db     823 RTAALNAVVSGAEER--------IADALDSRTMALDMTFSGAEEKIAEALDTRTAALGEL 874

Qy     310 ALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDS 369

Db     875 VASAETRIAG-ALDSRTDSLKTV---------------------VSGAEERIT-DVLDS 910

Qy     370 ATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTV 429

Db     911 RTMALDMSFSGVEEKITDILDGRTA------ALKSAVAGVEDRIAGALDSRTAALSG--- 961

Qy     430 KLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGIS 473

Db     962 -IVSGAEERIAEALDSRTLALDMTISGVEERIAEAMDARASSLS 1004


RESULT 5
Q8U8W1
ID   Q8U8W1      PRELIMINARY;      PRT;  1009 AA.
AC   Q8U8W1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Ice nucleation protein homolog.
GN   INA OR ATU3977 OR AGR_L_1758.
OS   Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=176299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608550; PubMed=11743193;
RA   Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA   Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA   Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA   Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA   Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA   Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA   Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA   Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA   Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA   Nester E.W.;
RT   "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT   C58.";
RL   Science 294:2317-2323(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21608551; PubMed=11743194;
```

```
RA   Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA   Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA   Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA   Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA   Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA   Cielo C., Slater S.;
RT   "Genome sequence of the plant pathogen and biotechnology agent
RT   Agrobacterium tumefaciens C58.";
RL   Science 294:2323-2328(2001).
DR   EMBL; AE009328; AAL44779.1; ALT_INIT.
DR   EMBL; AE008285; AAK89449.1; -.
KW   Complete proteome.
SQ   SEQUENCE   1009 AA;  100468 MW;  14D842830E8C65EE CRC64;

  Query Match             6.2%;  Score 149;  DB 16;  Length 1009;
  Best Local Similarity   22.4%;  Pred. No. 2.9;
  Matches  127;  Conservative   77;  Mismatches  228;  Indels  136;  Gaps   25;

Qy       7 AQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAI-TDYLKH---HVFAAHRFSVI 62
Db     118 ASQIDAMSTDQIKALNSSQVAGLSSAQVATLSSDELALFTTDEIKSISANAIAGLSAAAI 177

Qy      63 GSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKF------DRLATTA 116
Db     178 AGLSTDNAAALTKSQIAAMSSTQFNALTSGGSLATFSADEVKAISNKILAGLDVTKLSTGN 237

Qy     117 SSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARG 176
Db     238 IAALSKAQVSALSTTQFAAMSTDQ-------IKALTSDQVAGLSSA--QVATLSSD---- 284

Qy     177 DLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIA----------------------VQ 215
Db     285 ELALFSTDEIKAIGANAVA-GLSAAALAALTTDNAAALTKTQIAGLSSTQLNALTSANLA 343

Qy     216 TFSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVA-----NAGFGDRMLSVQSRDQL 270
Db     344 TFSA-DEIKAISTKALA-------GLDVTKLSTGNIAALTKTQAASLSSTQFAAMSTDQI 395

Qy     271 RG--GAFVLGMKDKEPKAALSEE-----TDWLDAYKAIKSASYSGAALNA---------- 313
Db     396 KALTSEQVAGLSSAQVATLSSDELALFSTDEISAFSANAVAGLSTAALAALTTGNATALT 455

Qy     314 GKRMAGL---PLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSA 370
Db     456 KTQIAGLSSTQLNALTSGSIATFSADEVKAISTK--ALAG-----LDVTKLSTGNVA-AL 507

Qy     371 TKAAVSQLSNLVGSVGVFAGWTT-----------AGLATDPAVKKAES----FIQDKVKS 415
Db     508 SKAQVSALSTT-----QFAAMSTDQIKALTSEQVAGLSSAQVATLSSDELALFSTDEIKA 562

Qy     416 ---------TASSTTSYVADQTVKLAKT-VKDMSGEAISSTGASLRSTVNNLRHRSAPE- 464
Db     563 IGANAVAGLSAAALAALTTDNASALTKTQIAGMSSTQINAL------TSANLATFSADEI 616

Qy     465 -------------ADIEEGGISAFSRSE 479
Db     617 KAITTKALGGLDVTKLSTGNIAALTKAQ 644


RESULT 6
Q92K98
ID   Q92K98      PRELIMINARY;      PRT;  2089 AA.
AC   Q92K98;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical transmembrane protein SMc00190.
GN   R01816 OR SMC00190.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR   EMBL; AL591788; CAC46395.1; -.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   2089 AA;  222522 MW;  1D6334385A31ADAD CRC64;

  Query Match             6.1%;  Score 146;  DB 16;  Length 2089;
  Best Local Similarity   18.9%;  Pred. No. 12;
  Matches  106;  Conservative   84;  Mismatches  208;  Indels  162;  Gaps   17;

Qy       3 INQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVI 62
Db     188 LRSAARSMAEVAM---RLAEPETNAADRVMTVGQAVRREVSAMNEGIERTIARATELEAL 244

Qy      63 GSPD------------------------ERDAALAHNEQIDALV----------------E 83
Db     245 VHSEVSALERSYSENELRVRTLVQELGLEREAIIGHSDRIRTAIAGAHTKLKDDLETASE 304

Qy      84 TRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDW 143
Db     305 DIASRIAVSGEAFASLIDTRAAA-----LTDKSDHALEN---------------------- 338

Qy     144 LATPLKPLT-PLISG-----------------ALSGAMDQVGTKMMDRARGDLHYLSTSP 185
Db     339 LSTMLTTRTDALLSGLTTAGVALSNEFDARLDALSDNLTQRGEQLLSQFETRASTLDANT 398

Qy     186 DKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGV 245
Db     399 EKLNAALNERARQLNETLIARTRDLNESLR----------------IGQQAISGGLDDVL 442

Qy     246 STAGGLVANAGFGDRMLSVQSRD------QLRGGAF-------------VLGMKDKEPKA 286
Db     443 SSLNSALDEKGASFRQSLKSSADDAIMDLDLRGGFFEEKLQTTVGQLASAFDERFHEFAS 502

Qy     287 ALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGL 346
Db     503 AFDKRASQLDT-KLMESLHRINETVSGGSEAIGGALDSSVD---KINSALSEQSLT---L 555

Qy     347 ALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGS-----------VGVFAGWTTA 394
Db     556 ATALG----------ATQDFIEETIGSRTSELSSLIGNAHNRIESVLSDKTGSLMGALTE 605

Qy     395 ---------GLATDP---AVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEA 442
Db     606 AQERIENGFGQRADALANALTTSERSLTDGLDSRTSAFIEGLQSAHARIEQTLTGSTDEI 665

Qy     443 ISSTGASLRSTVNNLRHRSA 462
Db     666 TSAIAASQHRLDNTLSERTA 685


RESULT 7
O85472
ID   O85472      PRELIMINARY;      PRT;  2055 AA.
AC   O85472;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Extracellular matrix binding protein (Fragment).
GN   EMB.
```

```
OS   Abiotrophia defectiva.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC   Aerococcaceae; Abiotrophia.
OX   NCBI_TaxID=46125;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NVS-47;
RX   MEDLINE=99081722; PubMed=9864195;
RA   Manganelli R., van de Rijn I.;
RT   "Characterization of emb, a gene encoding the major adhesin of
RT   Streptococcus defectivus.";
RL   Infect. Immun. 67:50-56(1999).
DR   EMBL; AF067776; AAD03320.1; -.
DR   TIGRFAMs; TIGR01168; YSIRK_signal; 1.
FT   NON_TER    2055   2055
SQ   SEQUENCE   2055 AA;  215642 MW;  9699C11DDE93E2FD CRC64;

  Query Match             6.0%;  Score 143.5;  DB 2;  Length 2055;
  Best Local Similarity   22.2%;  Pred. No. 16;
  Matches  123;  Conservative   78;  Mismatches  249;  Indels  105;  Gaps   23;

Qy       3 INQSAQQP------PGVAMESFRTASDASLA--SSSVRSVSTTSCRD--LQAI----TDY 48
           ||: :|:|          |:  ||| ||::|  :|:  : : || ||  | |:    |
Db     726 INEISQRPDLTREEKQAFMDQVRTARDAAMAKVASAANNQAVTSARDQGLNAVNNLPTPA 785

Qy      49 LKHHVFAAHRFSVIGSPDERDAALAHN------EQIDAL-----VETRANRLYSEGETPA 97
             |:     |   | : | :  |:  |       || |||       :| |    ::  | |
Db     786 AKYPEALGH---VRQAADAKRQAIRDNANLTAEEQADALRQVDAAQTAAEAAINQNHTNA 842

Qy      98 TIAETFAKAEKF-------DRLATTASSAFENTPFAAAASVLQYMQPAINKGDWL-----A 145
           |:|:  :   |          |    |: | |    |||      : |||  : |     |
Db     843 TLAKADSDGVKAINDINPQPRSKPAANQALEQV--AAAK-----RQAINNNNQLTDEEKA 895

Qy     146 TPLKPLTPLISGA---LSGAMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPA 202
             :: :   :: |   :  | |  |      | |     : :|     | |::    :
Db     896 QAIQQVDQALANAKTQVQAANDNNGVNQAKTA-GTTAINNINPQGTQKAQAIAAIEAAEQ 954

Qy     203 LGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQ---------GAVDFGVSTAGGLVA 253
             |  :     : |    | :  : | | |:: :|          || | ||:    |:
Db     955 AKRLELQGRNDLTTEERNNALADLTAKAQAAKDAVNQARNNTGVAGAKDNGVAQIQGINP 1014

Qy     254 NA-----------GFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWL-DAYKA 300
           |            |:    |:  :|          |:|  ||: :  |   ||  |
Db    1015 TAVVKPDARNAIDQAARDKEAEFQANTKL---------TDEEKAAAIKKVQDAARDAKAA 1065

Qy     301 IKSASYSG---AALNAGKRMAGLPLDVATDGLKAV-RSLVSATSLTKNGL---------A 347
           |  |  :|    |:| ||         |  :||:  |  ||    |  :           |
Db    1066 IDRAGSNGDVNNAVNQGK--------AAIQAIKALDDSQPSAKDTAKAAIQNAADAKKAA 1117

Qy     348 LAGGYAGVSKLQKMATKNITDSATKA-AVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAE 406
           :    |  : :  | | : | | || |    |      |           ::  |||:  :
Db    1118 ITANNALTQEEKAAAIKQVEDEAAKAQAAVDASRSKADVDRAKDQGLQKISDVPAVQPPK 1177

Qy     407 SFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEAD 466
             |   |:   : : : |  | :  |: :   :    |  |  :|:    :  |
Db    1178 LNAIAAVDQAATDKKAVINNDTT-LTQEEKEAAIRKVDEEAAKARQAINDATSNADVAAK 1236

Qy     467 IEEGGISAFSRSETP 481
           :|  :  :  :||
Db    1237 QAQGTQAINNVPQTP 1251


RESULT 8
Q9HPA2
ID   Q9HPA2      PRELIMINARY;      PRT;   536 AA.
AC   Q9HPA2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Htr17 transducer.
GN   HTR17 OR VNG1733G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR   EMBL; AE005078; AAG19968.1; -.
DR   HSSP; P02942; 1QU7.
DR   InterPro; IPR004089; Chmtaxis_transd.
DR   InterPro; IPR003660; HAMP.
DR   InterPro; IPR004090; Me_chemotaxis.
DR   Pfam; PF00015; MCPsignal; 1.
DR   PRINTS; PR00260; CHEMTRNSDUCR.
DR   SMART; SM00304; HAMP; 1.
DR   SMART; SM00283; MA; 1.
KW   Complete proteome.
SQ   SEQUENCE   536 AA;  56815 MW;  3B33D2FD8701EE04 CRC64;

  Query Match             5.9%;  Score 141;  DB 17;  Length 536;
  Best Local Similarity   23.2%;  Pred. No. 3.3;
  Matches  119;  Conservative   78;  Mismatches  197;  Indels  120;  Gaps   23;

Qy      12 GVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAA 71
             |  |::| |    || :  :  |: :|   :  : | ||     |||       |   |
Db      63 GPVMDAFST----SLLAVLIGVVAISSTGTISIVED-LKTSNARAHR----RRKDAEQAK 113

Qy      72 LAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVL 131
             :||     | :|   :  |      |:  :| : |  :  :| : :    |||
Db     114 RRADEQRAQAQEAKAEAQDAREEAETLAAKLQERAAERDAQLSKKASEYSSVMEKAAS-- 171

Qy     132 QYMQPAINKGDWLATPLKPLTPLI-SGALSGAMDQVG----------TKMMDRAR----- 175
                 ||       ||    | : | || :||          :|: | |
Db     172 ---------GD--------LTVRANSSSDSDAMTEVGQAFNEMIAAFEEMVARTREFAGD 214

Qy     176 -----GDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPA 230
                |:|      :  : :| ||:  |     |:                | || |
Db     215 VADNSRDIHASVDEIEAASEQVAESVQEISAGTDRE----------------HTRLADA 257

Qy     231 LASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSE 290
                |: |||:  |: :   ||:       :| |: :| |:         ||:::
Db     258 TDELQSLSGAVE-EVAASASQVAD-------VSDQAADRGRQGS-------DDAAAAVTQ 302

Qy     291 ETDWLDAYKAIKS-ASYSGAALNAGKRMAGLPLDVATD--GLKAVRSLVSATSLTKNGLA 347
             ||      :|: | |:    |      :|:  |  |: ||   | |: : : |     ::|
Db     303 MTDIQSETEAVVSEANALAAEAQEIRRVVDLINDI-TDQTHLLALNASIEAARADQDGDG 361

Qy     348 LAGGYAGVSKLQKMATKNITD----------SATKAAVSQLSNLVGSVGVFAGWTTAGLA 397
             |    | | :| :  |   ||          |:|:: |     :  :|   :|  |    |
Db     362 FAVVAAEVKELAE-ETNEATDEIETLVDNLLSSTQSVVDDSEQMAATVS--SGTETVENA 418

Qy     398 -----------TDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAISST 446
                      ||  | :|   | |   : |:|     || || : : ::|:| |  ||
Db     419 LTALEEIGDHVTD--VNEAVQQIDDATDAQAAS-----ADTTVSMIEEIEDIS-EQNSSE 470

Qy     447 GASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480
           ||: :     : ::|  :|: :|      |::|
Db     471 AASVSAAA---QQQTASLSDVTDGVTRLVDRADT 501


RESULT 9
Q8XSW6
ID   Q8XSW6      PRELIMINARY;      PRT;   642 AA.
AC   Q8XSW6;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
```

```
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Probable flagellar hook-associated protein 1.
GN   FLGK OR RSP0351 OR RS00754.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OG   Plasmid megaplasmid.
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646078; CAD17502.1; -.
DR   InterPro; IPR001444; Flag_bb_rod.
DR   Pfam; PF00460; flg_bb_rod; 1.
DR   PROSITE; PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
KW   Plasmid; Complete proteome.
SQ   SEQUENCE   642 AA;  65396 MW;  5B5D4CF4A6989313 CRC64;

  Query Match             5.8%;  Score 140.5;  DB 16;  Length 642;
  Best Local Similarity   20.9%;  Pred. No. 4.5;
  Matches  119;  Conservative   67;  Mismatches  216;  Indels  167;  Gaps   21;

Qy      12 GVAMESFRTASDASLA----SSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDE 67
           | |: ||  |:| |||    |:| | |  ::   ||              ||:| |
Db     108 GSALTSFFNAAD-SLASNPSSTSARQVFLSAASTLQT------------RFNAISG--- 150

Qy      68 RDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAA 127
           : |:|::  |:|  |:| | : |  :  | : :  ||||        ||
Db     151 QMASLSN--QVNTQVQTQVNSVNSTTQQIAALNDQIAKAE-------AASGG-------- 193

Qy     128 ASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMD-----QVGTKMMDRARGDLHYLS 182
           |||  :  |     :: |   |  ::    |    ||       :|:  |
Db     194 -------QPANDLRDQRDQLVQTLNQSIKASVVQTSDGQYNIYVGNGQA-LVQGNQSYQL 245

Qy     183 TSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVR-TVLAPALASRPSVQGAV 241
           |:    :|   :||   |||    : |  :       |   |  | ||  |   :  ||
Db     246 TTVASQYDPTQLSVGYKSPAGTVTIDDSQLGGGALGGLMQFRQNTLIPAQNSLGRLAAAV 305

Qy     242 DFGV-------------------STAGGLVA----NAGFGDRMLSVQSRD---------Q 269
           |                      ||||  ||    | | |    :: : :         :
Db     306 SADVNTQNKEGMDLNGKLGTDLFSTAGPSVAASSSNTGTGALTATITNANAGQGYDYQVK 365

Qy     270 LRGGAFVLG-MKDKEPKAALSEETDWLDA------------------------------ 297
           |||: :   |      :|     :|
Db     366 YSGGAYTVSHYPDGSGAVTVSSWPTTVDGVTLNLTGSMASGDSFLVRPTVNAASTMQTLT 425

Qy     298 --YKAIKSASYSGAALNAGKRMAGLP--LDVATDGLKAVRSLVSATSLTKNGLALAG--- 350
           | |: :|  |   :| |    |       :  |   |  | || ||| :| :|:|
Db     426 SDYHAVAAA--SPVVVNQGSNNTGSTSVASIGVDSTYAGSPLASAVSLTYSGGSLSGFPG 483

Qy     351 ----------GYAGVSKLQKMA------------------------TKNITDSATKA 373
                    |:| :   : |                          | | ||    :
Db     484 SVTVTVNGTATTYSGTAPYTQGATYSFNGIQMSLTGTPAANDTFTVSANTANSTDGHNAS 543

Qy     374 AVSQLSNLV----GSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTV 429
           | :|| |      |:  : :||:          :| : :  :    ||||:   :  |
Db     544 AFAQLRNATVLDNGTTSISSGWSNLVTQVGIQASQASANLTSQKALLASSTSQQQSVSGV 603

Qy     430 KL-AKTVKDMSGEAISSTGASLRSTVNNL 457
           |  : :  |  :      | :  | |:|
Db     604 NLDDEAMNLMKYQQAYQASAKVMQTANSL 632
```

```
RESULT 10
Q931N9
ID   Q931N9      PRELIMINARY;      PRT;  1983 AA.
AC   Q931N9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Mrp protein.
GN   MRP OR SAV1758.
OS   Staphylococcus aureus (strain Mu50 / ATCC 700699).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=158878;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA   Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA   Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA   Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA   Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA   Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
DR   EMBL; AP003363; BAB57920.1; -.
DR   TIGRFAMs; TIGR01168; YSIRK_signal; 1.
KW   Complete proteome.
SQ   SEQUENCE   1983 AA;  215786 MW;  573C6E6DFFFC7FBF CRC64;

  Query Match             5.8%;  Score 140.5;  DB 16;  Length 1983;
  Best Local Similarity   19.3%;  Pred. No. 23;
  Matches  105;  Conservative   94;  Mismatches  203;  Indels  143;  Gaps   21;

Qy      20 TASDASL----ASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHN 75
           ||:: ::      :: |  | | : : :||||    |                | ::|
Db     930 TAANTNILNANTNADVEQVKTNAIQGIQAITPATK------------VKTDAKNA----- 972

Qy      76 EQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQ 135
             ||   ||: | :::  :  ||: |  |  :  |:   ||           ||   | :
Db     973 --IDKSAETQHNTIFNNND--ATLEEQQAAQQLLDQAVATAKQNIN-----AADTNQEVA 1023

Qy     136 PAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVS 195
           | ::|      ::| | : : | :          : |:||  :  :: :|    :    :
Db    1024 QAKDQGTQNIVVIQPATQVKTDARNA--------VNDKAREAITNINATPGATREEKQEA 1075

Qy     196 VKRHSPALGRQVVDMGIAVQTFSALNVVRT-------VLAPALASRPSVQGA-------- 240
           : | :   | : |:|:   | : :| :|         : | :  : :  |
Db    1076 INRVNTLKNRALTDIGV-TSTTAMVNSIRDDAVNQIGAVQPHVTKKQTATGVLNDLATAK 1134

Qy     241 ---------------------VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAF---- 275
                               ||  ::||   : |     |     |    ||   |
Db    1135 KQEINQNTNATTEEKQVALNQVDQELATA---INNINQADTNAEVDQAQQLGTKAINAIQ 1191

Qy     276 --------VLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMA--------- 318
                   |   ::   | |:|      ||     |:|: :   ||  :: |
Db    1192 PNIVKKPAALAQINQHYNAKLAEINATPDATNDEKNAAIN--TLNQDRQQAIESIKQANT 1249

Qy     319 GLPLD----VATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMA--TKNITDSATK 372
             :|     || : : ||: |       :: :       | |:|  :     | | ||   :
Db    1250 NAEVDQAATVAENNIDAVQVDVVKKQAARDKIT-----AEVAKRIEAVKQTPNATDEEKQ 1304

Qy     373 AAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKA-ESFIQDKVKSTASSTTSYV----ADQ 427
           |||:|:: |                | |: :  ::    |:| :| :   :  :    |:
Db    1305 AAVNQINQL---------------KDQAINQINQNQTNDQVDTTTNQAVNAIDNVEAEV 1348

Qy     428 TVK------LAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSETP 481
           :|      : | ||:   :  :     || ||| |      |   |  :|  ::|  :::|
Db    1349 VIKPKAIADIEKAVKEKQQQIDN----SLDSTDNEKEVASQALAKEKEKALAAIDQAQTN 1404

Qy     482 FQLRR 486
             |: :
```

```
Db   1405 SQVNQ 1409


RESULT 11
Q99TB0
ID   Q99TB0      PRELIMINARY;      PRT;  2186 AA.
AC   Q99TB0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SA1577.
GN   SA1577.
OS   Staphylococcus aureus (strain N315).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Staphylococcus.
OX   NCBI_TaxID=158879;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21311952; PubMed=11418146;
RA   Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA   Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
RA   Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
RA   Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA   Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA   Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA   Ogasawara N., Hayashi H., Hiramatsu K.;
RT   "Whole genome sequencing of meticillin-resistant Staphylococcus
RT   aureus.";
RL   Lancet 357:1225-1240(2001).
DR   EMBL; AP003135; BAB42845.1; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR   TIGRFAMs; TIGR01168; YSIRK_signal; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   2186 AA;  238448 MW;  0CC84836F1868A4F CRC64;

  Query Match             5.8%;  Score 140.5;  DB 16;  Length 2186;
  Best Local Similarity   19.3%;  Pred. No. 27;
  Matches  105;  Conservative   94;  Mismatches  203;  Indels  143;  Gaps   21;

Qy      20 TASDASL----ASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHN 75
           ||:: ::     :: |  | | : : :||||   |               | ::|
Db     930 TAANTNILNANTNADVEQVKTNAIQGIQAITPATK-----------VKTDAKNA----- 972

Qy      76 EQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQ 135
             ||   ||: | :::  :  ||: |  |  :  |:   ||          ||   | :
Db     973 --IDKSAETQHNTIFNNND--ATLEEQQAAQQLLDQAVATAKQNIN-----AADTNQEVA 1023

Qy     136 PAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVS 195
           | ::|      ::| | : : | :          : |:||  :  :: :|    :    :
Db    1024 QAKDQGTQNIVVIQPATQVKTDARNA--------VNDKAREAITNINATPGATREEKQEA 1075

Qy     196 VKRHSPALGRQVVDMGIAVQTFSALNVVRT-------VLAPALASRPSVQGA-------- 240
           : | :    | : |:|:   | : :| :|         : | :  : :  |
Db    1076 INRVNTLKNRALTDIGV-TSTTAMVNSIRDDAVNQIGAVQPHVTKKQTATGVLNDLATAK 1134

Qy     241 ---------------------VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAF---- 275
                               ||  ::||   : |    |    |    ||   |
Db    1135 KQEINQNTNATTEEKQVALNQVDQELATA---INNINQADTNAEVDQAQQLGTKAINAIQ 1191

Qy     276 --------VLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMA--------- 318
                  |   ::   | |:|    ||    |:|: :   ||  :: |
Db    1192 PNIVKKPAALAQINQHYNAKLAEINATPDATNDEKNAAIN--TLNQDRQQAIESIKQANT 1249

Qy     319 GLPLD----VATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMA--TKNITDSATK 372
             :|    || : : ||:  |       :: :      | |:|  :    | | ||   :
Db    1250 NAEVDQAATVAENNIDAVQVDVVKKQAARDKIT-----AEVAKRIEAVKQTPNATDEEKQ 1304

Qy     373 AAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKA-ESFIQDKVKSTASSTTSYV----ADQ 427
           |||:|:: |                 | |: :  ::   |:| :| :   : :    |:
Db    1305 AAVNQINQL----------------KDQAINQINQNQTNDQVDTTTNQAVNAIDNVEAEV 1348

Qy     428 TVK------LAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAPSRSETP 481
             :|      : | ||:   :  :    || || |     |   |  :|  ::|  :::|
Db    1349 VIKPKAIADIEKAVKEKQQQIDN----SLDSTDNEKEVASQALAKEKEKALAAIDQAQTN 1404

Qy     482 FQLRR 486
             |: :
Db    1405 SQVNQ 1409


RESULT 12
Q9FCQ5
ID   Q9FCQ5      PRELIMINARY;      PRT;   503 AA.
AC   Q9FCQ5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Flagellin (Fragment).
GN   FLIC.
OS   Pseudomonas fragi.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=296;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIMB8542;
RA   Bellingham N.F., Morgan J.A.W., Saunders J.R., Winstanley C.;
RT   "Flagellin gene sequence phylogeny in the genus Pseudomonas.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ297534; CAC03723.1; -.
DR   InterPro; IPR001492; FlagellinN.
DR   InterPro; IPR001029; Flagellin_C.
DR   Pfam; PF00700; Flagellin_C; 1.
DR   Pfam; PF00669; Flagellin_N; 1.
DR   PRINTS; PR00207; FLAGELLIN.
DR   ProDom; PD000316; Flagellin_C; 2.
FT   NON_TER       1      1
FT   NON_TER     503    503
SQ   SEQUENCE   503 AA;  51085 MW;  AE42EEFB521BF7C1 CRC64;

  Query Match             5.8%;  Score 140;  DB 2;  Length 503;
  Best Local Similarity   24.3%;  Pred. No. 3.4;
  Matches   99;  Conservative   43;  Mismatches  135;  Indels  130;  Gaps   19;

Qy     147 PLKPLTPLISGA---------LSGAMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVK 197
           |:| :   |: |           ||| |  | :: | | :|   | : ||  |  |   :
Db      10 PVKGMAVTIANANNGSSIAQTAEGAM-QESTNILQRLR-ELALQSANGDKSADDRASLQQ 67

Qy     198 RHSPALG-------------RQVVDMGIAVQTF-------------------------- 217
             :  :|             | ::|      |:|
Db      68 EFTAKVGELTRISSTTTFGSRNLLDGSFQNQSFQVGADANQTISFGMSDISATGLKGSYG 127

Qy     218 --SALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAF 275
             ||   | |: |       | || |      | |  | |    |: |::   |
Db     128 EASAAGGVSTLSANV------VGGANDAATFKATGAAAFAAVDDQTLTINGTD------- 174

Qy     276 VLGMKDKE-------------------------PKAALSEETDWLDAYKAIKSASYSGA 309
           :  || :                         |  ||  :|:     |   | : |
Db     175 IAIAKDSKIADAVAEINKQTSKTGVTASADAATGTKLTLSSASDFTAVGSAASDAGFVAA 234

Qy     310 ALNAGKRMAGLPLDVATDGLK---AVRSLV-SATSLTK-NGLALAGGYAGVSKLQKMATK 364
           |  | | : | |:::  :|:|   |  ||  :| ::|| |  :  |  | |   : : |
Db     235 ATPA-KNLLGRPVEIQVNGVKVTIAAGSLENAAAAITKANTDSKTGVNASVKDGRLVLTS 293

Qy     365 ------NITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAV---------KKAESFI 409
                 |: ||         || |    |: || | | |  | :|         || :|
Db     294 EKGQAINLADSTGTGGPGSLSKL----GLTAGSTQAKLTNDTSVSFNGVGVKPKKGDS-- 347

Qy     410 QDKVKSTASS-----TTSYVADQTVKLAKTVKDMSGEAISSTGASLR 451
           | : |: :|      | |  || |:||  | ||      |:  | | |
Db     348 MDTIVSSINSASTGVTASKNADNTLKLFST-KD-----ITIAGGSCR 388


RESULT 13
```

```
O64046
ID   O64046      PRELIMINARY;      PRT;  2285 AA.
AC   O64046;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative transglycosylase.
GN   YOMI.
OS   Bacteriophage SPBc2.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX   NCBI_TaxID=66797;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Mauel C.,
RA   Karamata D.;
RT   "The complete nucleotide sequence of the Bacillus subtilis SPbetac2
RT   prophage.";
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF020713; AAC13005.1; -.
DR   HSSP; P00718; 153L.
DR   InterPro; IPR002886; Peptidase_M37.
DR   InterPro; IPR000189; SLT_domain.
DR   Pfam; PF01551; Peptidase_M37; 1.
DR   Pfam; PF01464; SLT; 1.
DR   PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
SQ   SEQUENCE   2285 AA;  252310 MW;  FF602C227754B357 CRC64;

  Query Match             5.8%;  Score 140;  DB 9;  Length 2285;
  Best Local Similarity   20.5%;  Pred. No. 30;
  Matches  128;  Conservative   80;  Mismatches  224;  Indels  192;  Gaps   26;

Qy      15 MESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSV-IGSPDERDAALA 73
           :| ::  :  :: : : |  |:    : ||: |||     | :  :|  || : |    :
Db     251 IELYQRQAQVNVQNLNTRYGSSMGSSNRQAVQDYLN----AVNSLNVSTGSNNIRSQIQS 306

Qy      74 HNEQIDAL---VETRANRLYSEGETPATIAETF----------------------AKA 106
           | |   |    :| ||:  | |   | : :||                      ::|
Db     307 LNMQFRELASNAQTAANQASSFG---AELTQTFKSMSTYLISGSLFYGAISGLKEMVSQA 363

Qy     107 EKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQV 166
           : | | |          :    :|   | :|: || |:  :  :  :          |:
Db     364 IEIDTLMTNIRRVMNEPDYKYNELL---QESIDLGDTLSNKITDILQMTGDFGRMGFDES 420

Qy     167 GTKMMDRARGDLHYLS-TSPDKLHDAMAVSVKRHSPALGRQV--------VDMGIAVQTF 217
           : :    |  :|  :||   : :  ::   : |    :          ||   || |
Db     421 ELSTLTKTAQVLQNVSDLTPDDTVNTLTAAMLNFNIAANDSISIADKLNEVDNNYAVTTL 480

Qy     218 SALNVVRTVLAPALASRPSVQGAVDFGVSTA--------GGLVAN------AGFGDRMLS 263
           | :|  | : ||   |:     | :||         | :| |      |  |:   |
Db     481 DLANSIRK--AGSTASTFGVELNDLIGYTTAIASTTRESGNIVGNSLKTIFARIGNNQSS 538

Qy     264 VQSRDQLRGGAFVLGMKDKEPKAALSE---ETDWL-DAYKAIKSASYSG---------- 308
           ::: :|:       | : |     :||   : | | || |   |   :|
Db     539 IKALEQIGISVKTAGGEAKSASDLISEVAGKWDTLSDAQKQNTSIGVAGIYQLSRFNAMM 598

Qy     309 ----AALNAGK---------------------------------------------- 315
               | || |
Db     599 NNFSIAQNAAKTAANSTGSAWSEQQKYADSLQARVNKLQNNFTEFAIAASDAFISDGLIE 658

Qy     316 --RMAGLPLDVATDGLKAV---RSLVSATS-----LTKNGLALAGG-YAGVSKL--QKMA 362
             : ||  |: :|  :|:|     |::| |      |:||   ||     |   :  : :|
Db     659 FTQAAGSLLNASTGVIKSVGFLPPLLAAVSTATLLLSKNTRTLASSLILGTRAMGQETLA 718

Qy     363 TKNITDSATKAAVSQ----------LSNLVGSVGVFAGWTTAGLATDPA-VKKA----- 405
           |  :    |:|||:           :| |||      ||    | :  |  |||
Db     719 TAGLEAGMTRAAVASRVLKTALRGLLVSTLVGGAFAALGWALESLISSFAEAKKAKDDFE 778

Qy     406 ----------------ESFIQ-----DKVKSTASSTTSYVADQ---TVKLAKT----VK 436
                          :  ||      ||| : | |:    :    | :||:|   ||
Db     779 QSQQTNVEAITTNKDSTDKLIQQYKELQKVKESRSLTSDEEQEYLQVTQQLAQTFPALVK 838

Qy     437 --DMSGEAISSTGASLRSTVNNLR 458
             |  | ||  |   |   : | :
```

```
Db     839 GYDSQGNAILKTNKELEKAIENTK 862


RESULT 14
O31976
ID   O31976      PRELIMINARY;      PRT;  2285 AA.
AC   O31976;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   YOMI protein.
GN   YOMI.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99115; CAB14053.1; -.
DR   HSSP; P00718; 153L.
DR   InterPro; IPR002886; Peptidase_M37.
DR   InterPro; IPR000189; SLT_domain.
DR   Pfam; PF01551; Peptidase_M37; 1.
DR   Pfam; PF01464; SLT; 1.
DR   PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
KW   Complete proteome.
SQ   SEQUENCE   2285 AA;  252310 MW;  FF602C227754B357 CRC64;

  Query Match             5.8%;  Score 140;  DB 16;  Length 2285;
  Best Local Similarity   20.5%;  Pred. No. 30;
  Matches  128;  Conservative   80;  Mismatches  224;  Indels  192;  Gaps   26;

Qy      15 MESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSV-IGSPDERDAALA 73
           :| ::  :  :: : : |  |:    : ||: |||     | :  :|  || : |    :
Db     251 IELYQRQAQVNVQNLNTRYGSSMGSSNRQAVQDYLN----AVNSLNVSTGSNNIRSQIQS 306
```

```
Qy      74 HNEQIDAL---VETRANRLYSEGETPATIAETF-----------------------AKA 106
           | |   |     :| ||:  | |   | : :||                       ::|
Db     307 LNMQFRELASNAQTAANQASSFG---AELTQTFKSMSTYLISGSLFYGAISGLKEMVSQA 363

Qy     107 EKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQV 166
           : | | |          :    :|   | :|: || |:  :  :  :          |:
Db     364 IEIDTLMTNIRRVMNEPDYKYNELL---QESIDLGDTLSNKITDILQMTGDFGRMGFDES 420

Qy     167 GTKMMDRARGDLHYLS-TSPDKLHDAMAVSVKRHSPALGRQV--------VDMGIAVQTF 217
           : :    |  :|  :||  : :  ::   : |    :           ||   || |
Db     421 ELSTLTKTAQVLQNVSDLTPDDTVNTLTAAMLNFNIAANDSISIADKLNEVDNNYAVTTL 480

Qy     218 SALNVVRTVLAPALASRPSVQGAVDFGVSTA--------GGLVAN------AGFGDRMLS 263
           | :|   | : ||   |:     | :||         | :| |       |  |:   |
Db     481 DLANSIRK--AGSTASTFGVELNDLIGYTTAIASTTRESGNIVGNSLKTIFARIGNNQSS 538

Qy     264 VQSRDQLRGGAFVLGMKDKEPKAALSE---ETDWL-DAYKAIKSASYSG---------- 308
           ::: :|:       | : |     :||   : | | || |    |   :|
Db     539 IKALEQIGISVKTAGGEAKSASDLISEVAGKWDTLSDAQKQNTSIGVAGIYQLSRFNAMM 598

Qy     309 ----AALNAGK---------------------------------------------- 315
               | || |
Db     599 NNFSIAQNAAKTAANSTGSAWSEQQKYADSLQARVNKLQNNFTEFAIAASDAFISDGLIE 658

Qy     316 --RMAGLPLDVATDGLKAV---RSLVSATS-----LTKNGLALAGG-YAGVSKL--QKMA 362
             : ||  |: :|  :|:|     |::| |     |:||  ||      |  :  : :|
Db     659 FTQAAGSLLNASTGVIKSVGFLPPLLAAVSTATLLLSKNTRTLASSLILGTRAMGQETLA 718

Qy     363 TKNITDSATKAAVSQ----------LSNLVGSVGVFAGWTTAGLATDPA-VKKA----- 405
           |  :    |:|||:           :| |||      ||    | :  |  |||
Db     719 TAGLEAGMTRAAVASRVLKTALRGLLVSTLVGGAFAALGWALESLISSFAEAKKAKDDFE 778

Qy     406 -----------------ESFIQ-----DKVKSTASSTTSYVADQ---TVKLAKT----VK 436
                           :  ||       |||  : | |:    :    | :||:|    ||
Db     779 QSQQTNVEAITTNKDSTDKLIQQYKELQKVKESRSLTSDEEQEYLQVTQQLAQTFPALVK 838

Qy     437 --DMSGEAISSTGASLRSTVNNLR 458
             |  | ||  |   |   : | :
Db     839 GYDSQGNAILKTNKELEKAIENTK 862
```

RESULT 15
Q93GT1
ID Q93GT1 PRELIMINARY; PRT; 491 AA.
AC Q93GT1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Flagellin (Fragment).
GN FLAA.
OS Campylobacter lari.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=201;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC12896;
RA Sekizuka T., Gondo T., Murayama O., Moore J.E., Matsuda M.;
RT "Genotypic and phenotypic characterization of flagellin from strains
RT of urease-positive thermophilic Campylobacter (UPTC).";
RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB073918; BAB71801.1; -.
DR InterPro; IPR001492; FlagellinN.
DR InterPro; IPR001029; Flagellin_C.
DR Pfam; PF00700; Flagellin_C; 1.
DR Pfam; PF00669; Flagellin_N; 1.
DR ProDom; PD000316; Flagellin_C; 2.
FT NON_TER 1 1
FT NON_TER 491 491
SQ SEQUENCE 491 AA; 50360 MW; 65080B0DB25FF3EE CRC64;

Query Match 5.8%; Score 139.5; DB 2; Length 491;

Best Local Similarity 23.8%; Pred. No. 3.5;
Matches 114; Conservative 66; Mismatches 183; Indels 117; Gaps 24;

```
Qy      21 ASDASLA--SSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAAL-----A 73
           | |:|||  || :|  |          || | ||     |       | : :: :: |     |
Db      23 ALDSSLARLSSGLRINSAADDASGLAIADSLKTQ--ANTLGQAINNANDANSMLQIADKA 80

Qy      74 HNEQIDAL--VETRANRLYSEGETPATIA----ETFAKAEKFDRLATTA--------SSA 119
           :||:   |  :: :| :   :|:|  | |     |      |: | :| |         | :
Db      81 MDEQLKILDTIKVKATQAAQDGQTAKTRAMIQGEINKLMEELDNIANTTTYNGKQLLSGS 140

Qy     120 FENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRAR---- 175
           | |  |              ||     : :  |   |  : |  |:     ||
Db     141 FSNQQFQI-------------GD---KANQTVNATIGSTNSAKIGQ--TRFETGARVTSG 182

Qy     176 GDLHYLSTSPDKLHDAMAVSVKRHS---PALGRQVVDMG-IAVQT-FSALNVVRTVLAPA 230
           ||:     : | ::|    ||   :    ||    ::  :: ||   |   |:|:   |
Db     183 GDIAITLKNYDGINDYKFQSVTVSTGVGTGLGALAAEINKVSDQTGVRASATVQTISTAA 242

Qy     231 LASRPSVQGAVDFGVSTAGGLVANAGFGDRML--SVQSRDQLRG-------GAFVLGMKD 281
           | :  : :|    ||   | :|  ||  |  |  :: ::    |       |  ||   |
Db     243 LTAGSTGEGFTINGV-VIGKVVFKAGDKDGALVSAINAKKDTTGVEASIVDGKLVLNSAD 301

Qy     282 ----------KEPKAALSEE---------TDWLDAYKAIKSASYSGAALNAGKRMAGLPL 322
                     :  | ::||           |  |   :   | :|  | |||        |
Db     302 GRGIKLSGSIGDLDAQIAEENYGRLTLVKNDGSDIVVSTTGAGFSTGATNAG-------L 354

Qy     323 DVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQ--------------KMATKNIT 367
             ||  |::::  :||       :| | |:  :|| |                ||  :|
Db     355 SEATVNLESIKGEISA------DIASAMGFNAMSKAQIGTGKQSACVTTLQGAMAVMDIA 408

Qy     368 DSATKAAVSQLSNLVGSVGVFAGWTTAGL----ATDPAVKKAESFIQDKVKSTASSTTSY 423
           |:    |:: |  :  ::|      |: :      |   || ||| |:|     || : :|
Db     409 DT----AIANLDTIRANIGATQNQITSTINNISVTQVNVKAAESQIRD--VDFASESANY 462
```

Search completed: January 2, 2003, 15:20:39
Job time : 37.0308 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:14:20 ; Search time 9.50976 Seconds
(without alignments)
2124.025 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINQSAQQPPGVAMESFRT.....EEGGISAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	138	5.7	1120	1	STFR_ECOLI	P76072 escherichia
2	137	5.7	540	1	CH60_LACHE	O68324 lactobacill
3	130	5.4	500	1	FLJB_SALAE	P52615 salmonella
4	130	5.4	1025	1	SLAP_CAUCR	P35828 caulobacter
5	129	5.4	436	1	Y868_CHLMU	Q9pjg1 chlamydia m
6	128.5	5.3	2388	1	SPCP_RAT	Q9qwn8 rattus norv
7	127	5.3	397	1	NCCB_ALCXX	Q44585 alcaligenes
8	123	5.1	544	1	CH60_BACHD	O50305 bacillus ha
9	121.5	5.0	1068	1	HIPR_MOUSE	Q9jky5 mus musculu
10	120.5	5.0	1140	1	YM96_YEAST	Q04893 saccharomyc
11	120	5.0	1306	1	MSB2_YEAST	P32334 saccharomyc
12	119.5	5.0	1608	1	HLYA_SERMA	P15320 serratia ma
13	119	4.9	657	1	HS7F_CAEEL	P11141 caenorhabdi
14	118.5	4.9	760	1	METE_MYCLE	O05564 mycobacteri
15	118.5	4.9	880	1	LYTD_BACSU	P39848 bacillus su
16	117	4.9	583	1	YP65_MYCTU	Q50733 mycobacteri
17	117	4.9	682	1	GR78_YEAST	P16474 saccharomyc
18	116.5	4.8	570	1	FLIF_RHOSH	Q53151 rhodobacter
19	116	4.8	763	1	HTR2_HALN1	Q9hp81 halobacteri
20	116	4.8	764	1	HTR2_HALSA	P71410 halobacteri
21	116	4.8	1577	1	HLYA_PROMI	P16466 proteus mir
22	116	4.8	3591	1	FHAB_BORPE	P12255 bordetella
23	115.5	4.8	401	1	YOPB_YERPS	Q06114 yersinia ps
24	115.5	4.8	1065	1	SED4_YEAST	P25365 saccharomyc
25	115	4.8	679	1	GR78_KLULA	P22010 kluyveromyc
26	115	4.8	1239	1	V120_EBV	P03189 epstein-bar
27	113.5	4.7	571	1	FLA1_CAMJE	P56963 campylobact
28	113	4.7	2541	1	TALI_HUMAN	Q9y490 homo sapien
29	112.5	4.7	505	1	FLJB_SALTY	P52616 salmonella
30	112.5	4.7	2090	1	N214_HUMAN	P35658 homo sapien
31	112	4.7	439	1	Y579_CHLTR	O84583 chlamydia t
32	112	4.7	492	1	FLIC_SALRU	P06175 salmonella
33	112	4.7	1027	1	AF10_HUMAN	P55197 homo sapien
34	111.5	4.6	547	1	CH60_ALTHA	Q9xau7 alteromonas
35	111.5	4.6	553	1	MCP2_ECOLI	P07017 escherichia
36	111	4.6	881	1	YJH8_YEAST	P47033 saccharomyc
37	110.5	4.6	543	1	CH60_MYCGE	P47632 mycoplasma
38	110.5	4.6	1419	1	ALA1_CANAL	O13368 candida alb
39	110	4.6	545	1	HTR5_HALN1	Q48318 halobacteri
40	110	4.6	814	1	SLA1_BACAN	P49051 bacillus an
41	109.5	4.5	901	1	SOK1_YEAST	P40317 saccharomyc
42	109.5	4.5	2541	1	TALI_MOUSE	P26039 mus musculu
43	109	4.5	1902	1	P3P_LACLC	P15292 lactococcus
44	109	4.5	2090	1	HFC1_MESAU	P51611 mesocricetu
45	108.5	4.5	465	1	SLAP_LACBR	Q05044 lactobacill

ALIGNMENTS

RESULT 1
STFR_ECOLI
ID STFR_ECOLI STANDARD; PRT; 1120 AA.
AC P76072; P77560;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Side tail fiber protein homolog from lambdoid prophage Rac.
GN STFR OR B1372.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -!- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AE000234; AAC74454.1; ALT_INIT.
DR EMBL; D90774; BAA14966.1; -.
DR EMBL; D90775; BAA14975.1; -.
DR EcoGene; EG13370; stfR.
DR InterPro; IPR004089; Chmtaxis_transd.
DR InterPro; IPR005003; Phage_fiber.
DR InterPro; IPR005068; Phage_fiber_2.
DR Pfam; PF03335; Phage_fiber; 6.
DR Pfam; PF03406; Phage_fiber_2; 1.
KW Hypothetical protein; Fiber protein; Repeat; Complete proteome.

```
SQ   SEQUENCE   1120 AA;  113779 MW;  542E59D71EE795B4 CRC64;

  Query Match             5.7%;  Score 138;  DB 1;  Length 1120;
  Best Local Similarity   21.1%;  Pred. No. 0.62;
  Matches  106;  Conservative   65;  Mismatches  221;  Indels  110;  Gaps   17;

Qy       3 INQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVI 62
           : : |:    ||  :      || ||:| |  :| :
Db     110 VEEVARNASAVAQNTAAAKKSASDASTSAREAATHAA---------------------- 146

Qy      63 GSPDERDAALAHNEQIDALVETRANRLYSEG--ETPATIA-ETFAKAEKFDRLATTASSA 119
             |  |:| | :        : :   | |   | || | :: | ||     | |:: |
Db     147 ---DAADSARAASTSAGQAASSAQSASSSAGTASTKATEASKSAAAAESSKSAAATSAGA 203

Qy     120 FENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLH 179
            : :   |:: ||           ||     |  | | : | |   :|  : |:
Db     204 AKTSETNASASLQS----------AATSASTATTKASEAATSARDAAASK--EAAKSSET 251

Qy     180 YLSTSPD------------KLHDAMAVSVKRHSPALGRQV-VDMGIAVQTFSALNVVRT 225
             |:|              |       : :    | |:      |      |: :   |
Db     252 NASSSASSAASSATAAGNSAKAAKTSETNARSSETAAGQSASAAAGSKTAAASSASAAST 311

Qy     226 VLAPALASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPK 285
           | ||  :  : :   |:|      ||     | :|       |  |
Db     312 SAGQASASATAAGKSAESAASSASTATTKAGEATEQASAAARS-------ASAAKTSETN 364

Qy     286 AALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNG 345
           |  ||  :   :: |   ::| | || :|     |    | ||    | :|  |||: :
Db     365 AKASETS--AESSKTAAASSASSAASSASS--ASASKDEATRQASAAKS--SATTASTKA 418

Qy     346 LALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKA 405
             ||                  |||  || |: :           :|    : | |:|
Db     419 TEAAG-----------------SATAAAQSKSTA----------ESAATRAETAAKRA 449

Qy     406 ESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPE- 464
           |          ||:|   :    |:|:      | | :::|  :::|  :|   |   :
Db     450 EDIASAVALEDASTTKKGI----VQLSSATNSTS-ETLAATPKAVKSAYDNAEKRLQKDQ 504

Qy     465 --ADIEEGG-----ISAFSRSE 479
             ||| : |     |:| |:::
Db     505 NGADIPDKGCFLNNINAVSKTD 526


RESULT 2
CH60_LACHE
ID   CH60_LACHE     STANDARD;      PRT;   540 AA.
AC   O68324;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR MOPA OR GROEL.
OS   Lactobacillus helveticus.
OC   Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=1587;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LH212;
RX   MEDLINE=98439363; PubMed=9766226;
RA   Broadbent J.R., Oberg C.J., Wei L.;
RT   "Characterization of the Lactobacillus helveticus groESL operon.";
RL   Res. Microbiol. 149:247-253(1998).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF031929; AAC29004.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
SQ   SEQUENCE   540 AA;  57638 MW;  4257DDB45FC7C4B4 CRC64;

  Query Match             5.7%;  Score 137;  DB 1;  Length 540;
  Best Local Similarity   20.8%;  Pred. No. 0.28;
  Matches  109;  Conservative   72;  Mismatches  198;  Indels  144;  Gaps   24;

Qy      21 ASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVI----GSPDERDAALAHNE 76
             | |    : :: ||:     : :  : | :|  :     |  |:     |:||  :  :
Db       2 AKDIKFSENARRSL----LKGVDKLADTVKTTIGPKGRNVVLEQSYGNPDITNDGVTIAK 57

Qy      77 QID------------ALVETRANRLYSEGETPATIAETFAKAEKFDRLATTAS-----S 118
             |:            |    : | :  :| | ||:       |    :   |:
Db      58 SIELKDRYENMGAKLVAEAAQKTNDIAGDGTTTATVLTQAIAREGMKNVTAGANPVGIRR 117

Qy     119 AFENTPFAAASVLQYMQPAINKGDWLA--TPLKPLTPLISGALSGAMDQVG--------- 167
             |     ||   | :   :   | :|    :   :  |   :: ||:||
Db     118 GIEKATKAAVDELHKISHKVESKDQIANVAAVSSASKEIGALIADAMEKVGHDGVITIED 177

Qy     168 ----------TKMMDRARGDL-HYLSTSPDKLH----------DAMAVSVKRHSPALGR 205
                     : |   || |  |: |  ||:           |    :::   | | :
Db     178 SRGINTELSVVEGMQFDRGYLSQYMVTDNDKMEADLDNPYILITDKKISNIQDILPLL-Q 236

Qy     206 QVVDMGIAVQTFSALNVVRTVLAPALAS--RPSVQGAVDFGVSTAGGLVANAGFGD-RML 262
           ::|  |      | | :   :   || :     ::|   | |       |   |||| |
Db     237 EIVQQG-----KSLLIIADDITGEALPTLVLNKIRGT--FNVVA----VKAPGFGDRRKA 285

Qy     263 SVQSRDQLRGGAFV---LGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAG 319
             :|    | ||  :   ||:: |: |
Db     286 QLQDIAALTGGTVITEDLGLELKDTK---------------------------------- 311

Qy     320 LPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAG--------VSKLQKMATKNITDSAT 371
                 | |   | :    ::||:   : || ||         |  ::|    : :|
Db     312 ------IDQLGQARRI----TVTKDSTTIVGG-AGSKEAIDERVDTIRKQIEDSTSDFDK 360

Qy     372 KAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKST-ASSTTSYVA---DQ 427
           |    :|: | | | |       | ||:  :|:    |:| : || |:    |||
Db     361 KKLQERLAKLTGGVAVI----HVGAATETELKERRYRIEDALNSTRAAVDEGYVAGGGTA 416

Qy     428 TVKLAKTVKDMSGEAI-SSTGASLRSTVNNLRHRSAPEADIEE 469
             | : | |::: ||     || ::      ||  |||   | |
Db     417 LVNVEKAVREVKGETTDEQTGINI-----VLRALSAPVRQIAE 454


RESULT 3
FLJB_SALAE
ID   FLJB_SALAE     STANDARD;      PRT;   500 AA.
AC   P52615;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Phase-2 flagellin.
GN   FLJB.
OS   Salmonella abortus-equi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=607;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93139045; PubMed=8423149;
```

```
RA   Okazaki N., Matsuo S., Saito K., Tominaga A., Enomoto M.;
RT   "Conversion of the Salmonella phase 1 flagellin gene fliC to the
RT   phase 2 gene fljB on the Escherichia coli K-12 chromosome.";
RL   J. Bacteriol. 175:758-766(1993).
RN   [2]
RP   SEQUENCE OF 478-501 FROM N.A.
RX   MEDLINE=93173100; PubMed=8437573;
RA   Hanafusa T., Saitko K., Tominaga A., Enomoto M.;
RT   "Nucleotide sequence and regulated expression of the Salmonella fljA
RT   gene encoding a repressor of the phase 1 flagellin gene.";
RL   Mol. Gen. Genet. 236:260-266(1993).
CC   -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC       FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC   -!- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC       BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC       PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13690; BAA02848.1; -.
DR   EMBL; D12510; BAA02072.1; -.
DR   InterPro; IPR001492; FlagellinN.
DR   InterPro; IPR001029; Flagellin_C.
DR   Pfam; PF00669; Flagellin_N; 1.
DR   Pfam; PF00700; Flagellin_C; 1.
DR   PRINTS; PR00207; FLAGELLIN.
DR   ProDom; PD000316; Flagellin_C; 1.
KW   Flagella.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CONFLICT    493    493       Q -> L (IN REF. 2).
SQ   SEQUENCE   500 AA;  51927 MW;  336D6F8F0EAA56CC CRC64;

  Query Match             5.4%;  Score 130;  DB 1;  Length 500;
  Best Local Similarity   21.9%;  Pred. No. 0.67;
  Matches  114;  Conservative   71;  Mismatches  232;  Indels  104;  Gaps   24;

Qy       2 HINQSAQQPPGVAME------SFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFA 55
           ::|:| |   | |:|         :| |   |  :: :  | : : |   :      :
Db      16 NLNKS-QSALGTAIERLSSGLRINSAKD-DAAGQAIANRFTANIKGLTQASRNANDGISI 73

Qy      56 AHRFSVIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATT 115
           ||  :  |: :| :  |   ::: |    ||   |: :  :  ||   :  : ||:  :
Db      74 AH--TTEGALNEINNNL---QRVRELAVQSANSTNSQSDLDSIQAEITQRLNEIDRV--S 126

Qy     116 ASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGA----LSGAMDQVGTKMM 171
             : |      |     :|   | |: :   || :     |     :  | |   | ::
Db     127 GQTQFNGVKVLAQDNTLTIQVGANDGETIDIDLKQINSQTLGLDSLNVQKAYDVSATDVI 186

Qy     172 DRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVV--DMGIAVQTFSALNVVRTVLAP 229
             |     |:|       |  :|    |||   |  |   |  :|       ||
Db     187 SSTYSDGTQALTAP------TATDIK---AALGNPTVTGDTLTAAVSFKDGKYYATVSGY 237

Qy     230 ALASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALS 289
             |   : |  :  | :| | |:   ||    :  ::  : |   |  ::   | ||
Db     238 TDAGDTAKNGKYEVTVDSATGAVS---FG----ATPTKSTVTGDTAVTKVQVNAPVAA-- 288

Qy     290 EETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALA 349
             ||  | | |   |  :|     |              :||  :   |||  :
Db     289 ------DA--ATKKALQDGGVSSADASAA--------------TLVKMSYTDKNGKTIE 325

Qy     350 GGYA---------------GVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTT 393
           ||||                |  |  |  :    |  || | :||  : |   |    |
Db     326 GGYALKAGDKYYAADYDEATGAIK-AKTTSYTAADGTTKTAANQLGGVDGKTEVV---TI 381

Qy     394 AGLATDPAVKKAESF-IQDKVKSTASSTTSYVADQ-TVKLAK---------TVKDMSGEA 442
             |   : :      |  | ::   |: ||     :    ||:          |::    |
Db     382 DGKTYNASKAAGHDFKAQPELAEAAAKTTENPLQKIDAALAQVDALRSDLGAVQNRFNSA 441

Qy     443 ISSTGASLRSTVNNLRHRSAPEADIEEGG----ISAFSRSE 479
           |:: |    :|||||   |   : ||:      :|  ||::
Db     442 ITNLG----NTVNNL---SEARSRIEDSDYATEVSNMSRAQ 475


RESULT 4
SLAP_CAUCR
ID   SLAP_CAUCR     STANDARD;      PRT;  1025 AA.
AC   P35828; Q46015; Q9RFI2;
DT   01-JUN-1994 (Rel. 29, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   S-layer protein (Paracrystalline surface layer protein).
GN   RSAA OR CC1007.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=93007489; PubMed=1393820;
RA   Gilchrist A., Fisher J.A., Smit J.K.;
RT   "Nucleotide sequence analysis of the gene encoding the Caulobacter
RT   crescentus paracrystalline surface layer protein.";
RL   Can. J. Microbiol. 38:193-202(1992).
RN   [2]
RP   REVISIONS TO 376; 636 AND 842-843.
RA   Awram P.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JS3001;
RA   Bingle W.H., Awram P.A., Nomellini J.F., Smit J.K.;
RT   "The secretion signal of C. crescentus S-layer protein is located in
RT   the C-terminal 82 amino acids of the molecule.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RN   [5]
RP   SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=89008089; PubMed=3049545;
RA   Fisher J.A., Smit J.K., Agabian N.;
RT   "Transcriptional analysis of the major surface array gene of
RT   Caulobacter crescentus.";
RL   J. Bacteriol. 170:4706-4713(1988).
RN   [6]
RP   CHARACTERIZATION.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=98292737; PubMed=9620954;
RA   Awram P., Smit J.K.;
RT   "The Caulobacter crescentus paracrystalline S-layer protein is
RT   secreted by an ABC transporter (type I) secretion apparatus.";
RL   J. Bacteriol. 180:3062-3069(1998).
CC   -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC       OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A
CC       PHYSICAL BARRIER TO PARASITES AND LYTIC ENZYMES.
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER
```

```
CC       (TYPE I) SECRETION APPARATUS.
CC    -!- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE
CC        SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A
CC        SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC    ---------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------
DR    EMBL; AF062345; AAC38665.2; -.
DR    EMBL; AF193063; AAF19365.1; -.
DR    EMBL; AE005779; AAK22991.1; ALT_INIT.
DR    HSSP; P22629; 1SWC.
DR    TIGR; CC1007; -.
DR    InterPro; IPR001343; Hemlysn_Ca_bind.
DR    Pfam; PF00353; hemolysinCabind; 3.
DR    PRINTS; PR00313; CABNDNGRPT.
KW    Cell wall; S-layer; Calcium-binding; Complete proteome.
FT    INIT_MET      0      0
SQ    SEQUENCE   1025 AA;  98001 MW;  AD7A326E1363D8AC CRC64;

  Query Match             5.4%;  Score 130;  DB 1;  Length 1025;
  Best Local Similarity   21.0%;  Pred. No. 1.7;
  Matches  101;  Conservative   75;  Mismatches  166;  Indels  140;  Gaps   23;

Qy      22 SDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDAL 81
           |||:  ::::: |::|:   :|       :  |     |  |             :| |
Db      41 SDAAALTNTLKLVNSTTAVAIQT------YQFFTGVAPSAAG-------------LDFL 80

Qy      82 VETRAN------RLYSEGETPATIAETFAKAEKF----DRLATTASSAFENTPFAAA--- 128
           |::  |       ||:          ||:  :|      |||  | :    | ||||
Db      81 VDSTTNTNDLNDAYYSK----------FAQENRFINFSINLATGAGAG--ATAFAAAYTG 128

Qy     129 -SVLQYMQPAINK--GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSP 185
            | | :  | :|  |: :||            :  :|         : ::||:
Db     129 VSYAQTVATAYDKIIGNAVAT------------AAGVDVAAAVAFLSRQANIDYLT--- 172

Qy     186 DKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVL---APALASRPSVQGAVD 242
                  |: ::|      :|: :       :    ||     | |:   :    |
Db     173 --------AFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAAMINDLSD 224

Qy     243 FGVST--AGGL-----VANAGFGDRMLSVQS-RDQLRG------------GAFVLGMKD 281
             :||  | |:        ::|     ||: :  | | |             ||  | : |
Db     225 GALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVAGEVAGAATLTVGD 284

Qy     282 KEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSL 341
           | ::  :|: |            ||: |      ||  |   |:: : :: |  ::
Db     285 TLSGGAGTDVLNWVQA----------AAVTA------LPTGVTISGIETM-NVTSGAAI 326

Qy     342 TKNGLALAGGYAGVSKL----------------QKMATKNITDSATKAAVSQLSNL-VGS 384
           | |    : |  |:: |              | :       :|   ||   :|: | |
Db     327 TLN---TSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGGANVTVAS 383

Qy     385 VGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAIS 444
           || :| || |     |   |   :   | :::::||  :| |   | ||   :| |::
Db     384 TGVTSGTTTVG-----ANSAASGTVSVSVANSSTTTTGAIA-VTGGTAVTVAQTAGNAVN 437

Qy     445 ST 446
           :|
Db     438 TT 439


RESULT 5
Y868_CHLMU
ID   Y868_CHLMU     STANDARD;      PRT;   436 AA.
AC   Q9PJG1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein TC0868.
GN   TC0868.
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MoPn / Nigg;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
CC   -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0808/CT579/TC0868
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE002353; AAF39664.1; -.
DR   TIGR; TC0868; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   436 AA;  43891 MW;  43E21C9924CD7A48 CRC64;

  Query Match             5.4%;  Score 129;  DB 1;  Length 436;
  Best Local Similarity   20.1%;  Pred. No. 0.64;
  Matches   85;  Conservative   67;  Mismatches  170;  Indels  100;  Gaps   14;

Qy      66 DERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRL-----ATTASSAF 120
           |:  ||    :| :|  |          |::        :|: | ::       || :|
Db      30 DKNGAAAKTQQQEEASFEDLIQDSQGTGKSKKASTSQTSKSGKSEKAQKSSGTTTTTSVA 89

Qy     121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGAL----SGAMDQVGTKMMDRARG 176
           : :  | |  ::  : :    |  |:    |  | ::| :      | :  :|  |   |:
Db      90 QASQTATAQAVRGARSSGFNNDGSASLPSPSNPNVNGVVLKKNMGTLALMGLIMTLLAQA 149

Qy     177 DLHYLSTSPDKLHDAMAVSVKRHSPALGRQVV---DMGIAVQTFSALNVVRTVLAPALAS 233
                 |:            | :: : |:  ||    ::| |::| :      | |    :
Db     150 SAKSWSS----------SFQQQNQAIQNQVAMAPEIGNAIRTQANHQAAATELQAKQSL 198

Qy     234 RPSVQGAVDFGVSTAGGLV-ANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEET 292
             :  | | ||  ||:: |:  |                |:|     || : ||
Db     199 ISGITNIVGFAVSVGGGILSASKSLG------------------GLK----SAAFTNET 235

Qy     293 DWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGY 352
                   |: : :| :|  : |   ||  :        |:||: ||    | :|
Db     236 ----------ATAASSATSAAAKTAVNALDDVAN--------VAATAGTK---AASGAA 273

Qy     353 AGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDK 412
           :  |      |:|:||:|:|      |   |  |:|
Db     274 SAASSAATKLTQNMTDTASKTLSQTASKATG--GLF------------------------ 307

Qy     413 VKSTASSTTSYVADQTVKLAKTVKDMSGEAISSTGASLRS--TVNNLRHRSAPEADIEEG 470
               | :| :: :::  :    ||   | |    | :| |  ::: : |       |   |
Db     308 --GNALNTPNW-SEKISRGLNVVKTQGGRAAQFAGRALSSAMSISQMVHGLTAGVDGITG 364

Qy     471 GI 472
           ||
Db     365 GI 366


RESULT 6
SPCP_RAT
ID   SPCP_RAT       STANDARD;      PRT;  2388 AA.
```

```
AC   Q9QWN8; O88197; Q9ES68;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Spectrin beta chain, brain 2 (Spectrin, non-erythroid beta chain 2)
DE   (Beta-III spectrin) (SPNB-3) (Beta SpIII sigma 1) (Spectrin-like
DE   protein GTRAP41).
GN   SPTBN2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=98340016; PubMed=9675416;
RA   Ohara O., Ohara R., Yamakawa H., Nakajima D., Nakayama M.;
RT   "Characterization of a new beta-spectrin gene which is predominantly
RT   expressed in brain.";
RL   Brain Res. Mol. Brain Res. 57:181-192(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=98369639; PubMed=9704016;
RA   Sakaguchi G., Orita S., Naito A., Maeda M., Igarashi H., Sasaki T.,
RA   Takai Y.;
RT   "A novel brain-specific isoform of beta spectrin: isolation and its
RT   interaction with Munc13.";
RL   Biochem. Biophys. Res. Commun. 248:846-851(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Rothstein J.D., Jackson M.;
RT   "GTRAP41- a cytoskeletal modulator of EAAT4 glutamate transporter.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probably plays an important role in neuronal membrane
CC       skeleton.
CC   -!- TISSUE SPECIFICITY: Abundantly transcribed in the brain. Neurons
CC       are the predominant cell-type to express the gene. Found
CC       abundantly in Purkinje cells.
CC   -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC   -!- SIMILARITY: CONTAINS 17 SPECTRIN REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB008551; BAA32699.1; -.
DR   EMBL; AB001347; BAA32473.1; -.
DR   EMBL; AF225960; AAG28596.1; -.
DR   HSSP; Q01082; 1BKR.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR002017; Spectrin.
DR   InterPro; IPR001605; Spectrin_PH.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00435; spectrin; 17.
DR   PRINTS; PR00683; SPECTRINPH.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00150; SPEC; 17.
DR   PROSITE; PS00019; ACTININ_1; 1.
DR   PROSITE; PS00020; ACTININ_2; 1.
DR   PROSITE; PS50021; CH; 2.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   Cytoskeleton; Membrane; Repeat; Actin-binding; Capping protein.
FT   DOMAIN       1    278      ACTIN-BINDING (BY SIMILARITY).
FT   DOMAIN      57    161      CH 1.
FT   DOMAIN     176    278      CH 2.
FT   REPEAT     305    415      SPECTRIN 1.
FT   REPEAT     425    529      SPECTRIN 2.
FT   REPEAT     531    639      SPECTRIN 3.
FT   REPEAT     641    745      SPECTRIN 4.
FT   REPEAT     747    850      SPECTRIN 5.
FT   REPEAT     852    956      SPECTRIN 6.
FT   REPEAT     958   1063      SPECTRIN 7.
FT   REPEAT    1065   1170      SPECTRIN 8.
FT   REPEAT    1172   1276      SPECTRIN 9.
FT   REPEAT    1278   1381      SPECTRIN 10.
FT   REPEAT    1383   1486      SPECTRIN 11.
FT   REPEAT    1488   1586      SPECTRIN 12.
FT   REPEAT    1588   1692      SPECTRIN 13.
FT   REPEAT    1694   1799      SPECTRIN 14.
FT   REPEAT    1801   1905      SPECTRIN 15.
FT   REPEAT    1907   2011      SPECTRIN 16.
FT   REPEAT    2013   2075      SPECTRIN 17.
FT   DOMAIN    2218   2328      PH.
FT   CONFLICT   326    328      VTL -> GTF (IN REF. 3).
FT   CONFLICT   543    543      L -> F (IN REF. 2).
FT   CONFLICT   608    608      D -> G (IN REF. 3).
FT   CONFLICT   887    887      L -> P (IN REF. 3).
FT   CONFLICT   908    908      V -> I (IN REF. 3).
FT   CONFLICT   948    948      D -> G (IN REF. 3).
FT   CONFLICT  1156   1157      EL -> GA (IN REF. 2).
FT   CONFLICT  1194   1194      F -> V (IN REF. 2 AND 3).
FT   CONFLICT  1555   1555      G -> R (IN REF. 3).
FT   CONFLICT  1769   1769      R -> W (IN REF. 3).
SQ   SEQUENCE   2388 AA;  271062 MW;  3EC8966AF0665F19 CRC64;

  Query Match             5.3%;  Score 128.5;  DB 1;  Length 2388;
  Best Local Similarity   21.8%;  Pred. No. 6.1;
  Matches  101;  Conservative   51;  Mismatches  168;  Indels  143;  Gaps   18;

Qy      35 STTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDA-------LVETRAN 87
           |  : |||  :   |  |   | |  : |        |   :|: |        |||
Db     666 SAETGRDLTGVLRLLNKH--TALRGEMSGRLGPLKLTLEQGQQLVAEGHPGANQASTRAA 723

Qy      88 RLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATP 147
           | :: |    :||      |:  |||              |||: |:   | :   ||
Db     724 ELQAQWERLEALAE-----ERAQRLA-----------QAASLYQFQADANDMEAWLVDA 766

Qy     148 LKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQV 207
           |:    |:|    |  |:  |: : |               | |:   :: | |
Db     767 LR----LVSSPEVG-HDEFSTQALARQ-------------HRALEEEIRAHRP------ 801

Qy     208 VDMGIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVANAGFGDR------- 260
           |  ||      | |||:  | |||  |       | | ||   |
Db     802 --------TLDALREQAAALPPALSHTPEVQGRVPTLEQHYEELQARAGERARALEAALA 853

Qy     261 ---MLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRM 317
             |||          ||  | :::||          ||:
Db     854 FYTMLS-------EAGACGLWVEEKE---------QWLN---------------------- 876

Qy     318 AGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSATKAAVSQ 377
           || |    : |: |:           | ||           :  |  :| :   ||:
Db     877 -GLALPERLEDLEVVQQRFETLEPEMNALA----------ARVTAVSDIAEQLLKASPPG 925

Qy     378 LSNLVGS----------VGVFAGWTTAGLATDPAVK-------KAESFIQDKVKSTASS- 419
             ::|:             |    | | :  :::       : :::::| |   |:
Db     926 KDRIIGTQEQLNQRWQQFRSLADGKKAALTSALSIQNYHLECTETQAWMREKTKVIESTQ 985

Qy     420 -----TTSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNL 457
                :| |  ||| | :|:  ||||:    |    | |
Db     986 DLGNDLAGVLALQR-KLAGTERDL--EAISARVGELTQEANAL 1025


RESULT 7
NCCB_ALCXX
ID   NCCB_ALCXX     STANDARD;      PRT;   397 AA.
```

```
AC   Q44585;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Nickel-cobalt-cadmium resistance protein nccB.
GN   NCCB.
OS   Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxidans).
OG   Plasmid pTOM9.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Achromobacter.
OX   NCBI_TaxID=515;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=31A;
RX   MEDLINE=95050278; PubMed=7961470;
RA   Schmidt T., Schlegel H.G.;
RT   "Combined nickel-cobalt-cadmium resistance encoded by the ncc locus
RT   of Alcaligenes xylosoxidans 31A.";
RL   J. Bacteriol. 176:7045-7054(1994).
CC   -!- FUNCTION: COMPONENT OF THE NCC CATION-EFFLUX SYSTEM THAT CONFERS
CC       RESISTANCE TO NICKEL, COBALT AND CADMIUM.
CC   -!- SIMILARITY: HIGH, TO A.EUTROPHUS CNRB AND SOME, TO ALCALIGENES
CC       CZCB.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L31363; AAA65105.1; -.
DR   InterPro; IPR002215; HlyD.
DR   Pfam; PF00529; HlyD; 1.
DR   TIGRFAMs; TIGR00999; 8a0102; 1.
KW   Plasmid; Nickel; Cobalt; Cadmium; Transport.
SQ   SEQUENCE   397 AA;  39933 MW;  C6DE57CE314996B8 CRC64;

  Query Match             5.3%;  Score 127;  DB 1;  Length 397;
  Best Local Similarity   22.8%;  Pred. No. 0.75;
  Matches   89;  Conservative   59;  Mismatches  140;  Indels  102;  Gaps   19;

Qy    149 KPLTPLISGALSGAMDQVGTKMMDRARG-------DLHYLSTSPDKLHDAMAVSVKRHSP 201
          :|  |:|:| :: |  ||      |||       ::  |: :|:|     | |      |
Db      9 QPSWPMIAG-VAAAAALVG---FGAARGLGSPSGAEVSKLAAAPEK----AAASAPAAEP 60

Qy    202 A----LGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVANAGF 257
          |    |  :    |||:  ||  |   :|||  ||  :| |:     | | | |
Db     61 AEVRIPGEYLAAANIAVEPVSAGGVGSVLLAP--ASVAAVPGSEAVIASRAAGAV----- 113

Qy    258 GDRMLSVQSR--DQLRGGAFVLGMKDKEPKAALSEETDWLDA--------YKAIKSASYS 307
          | :| :  | :| |  || : |    ||:: |     |         |:   |
Db    114 ----LRIQRKLGDAVRAGD-VLALVDSPEAAAMAAERKVAQARADLARKTYERESSLFQQ 168

Qy    308 GAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNG--LALAGGYAGVSKLQKMATKN 365
          |         | : ||||   ::   ::  |  :: :|  :|:    ||     ::  ::
Db    169 GVTPRQEMESARIALDVAQAEVQRAATVAQAAKVSSDGRSVAVVSPIAG-----RITAQS 223

Qy    366 ITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVA 425
          :|  |  |  ::|  : ||  |                 : |:::      ||: |:   |
Db    224 VTLGAYVAPQAELFRVAGSGAV----------------QVEAYV------TAADTSRIAA 261

Qy    426 --DQTVKLAK----------TVKDMSGEAISST---------------GASLR--STVN 455
            | |: ||                :|| | ::|                 |  :|  :
Db    262 GSDATIVLANGAPLAGRVQAVTPTVSGSARAATVVVTPVDANSGLIVGEGVQVRLHTKAA 321

Qy    456 NLRHRSAPEADIE--EGGISAFSRSETPFQ 483
          :    | ||  ::  :|    | |::  |:
Db    322 DANAMSVPEDAVQNLDGRDVVFVRTQQGFR 351


RESULT 8
```

```
CH60_BACHD
ID   CH60_BACHD     STANDARD;      PRT;   544 AA.
AC   O50305; Q9KFC3;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN   GROL OR MOPA OR GROEL OR BH0562.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=97141316; PubMed=8987660;
RA   Xu Y., Kobayashi T., Kudo T.;
RT   "Molecular cloning and nucleotide sequence of the groEL gene from the
RT   alkaliphilic Bacillus sp. strain C-125 and reactivation of thermally
RT   inactivated alpha-glucosidase by recombinant GroEL.";
RL   Biosci. Biotechnol. Biochem. 60:1633-1636(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- INDUCTION: BY HEAT SHOCK.
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D55630; BAA09494.1; -.
DR   EMBL; AP001508; BAB04281.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Heat shock; Complete proteome.
FT   CONFLICT    407    407       E -> VK (IN REF. 1).
FT   CONFLICT    421    421       MISSING (IN REF. 1).
FT   CONFLICT    500    500       A -> VP (IN REF. 1).
FT   CONFLICT    512    512       MISSING (IN REF. 1).
SQ   SEQUENCE   544 AA;  57403 MW;  C769ED81C903C971 CRC64;

  Query Match             5.1%;  Score 123;  DB 1;  Length 544;
  Best Local Similarity   18.9%;  Pred. No. 2;
  Matches  104;  Conservative   75;  Mismatches  192;  Indels  180;  Gaps   23;

Qy     21 ASDASLASSSVRSVSTTSCRDLQAITDYLK-------HHVFAAHRFSVIGSP-------- 65
          | |    :  : ||:     | :  : | :|        :|    :|   |||
Db      2 AKDIKFSEDARRSM----LRGVDKLADAVKVTLGPKGRNVVLEKKF---GSPLITNDGVT 54

Qy     66 -----DERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF 120
               :  ||      :: | | :: | :  :| | ||:    |:|   : |    | |
Db     55 IAKEIELEDAFENMGAKLVAEVASKTNDIAGDGTTTATV---LAQAMIREGLKNVTSGA- 110
```

```
Qy    121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLIS-GALSGAMDQVG---TKMMDRARG 176
Db    111 -NPMVIRKGIEKATQVAVEELSKISKPIEGKDSIAQVAAISSADDEVGKIIAEAMERVGN 169

Qy    177 D---------------------------LHYLSTSPDKLH-----------DAMAVSVK 197
Db    170 DGVITIEESKGFSTELEVVEGMQFDRGYASPYMVTDSDKMEAVLDNPYVLITDKKISNIQ 229

Qy    198 RHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVAN--- 254
Db    230 EVLPVL-EQVVQQGKPI-----LIIAEDVEGEALAT-----------------LVVNKLR 266

Qy    255 ----------AGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKAIKSA 304
Db    267 GTFNAVAVKAPGFGDR-----------------------RKAMLEDI------------ 290

Qy    305 SYSGAALNAGKRMA-GLPLDVATDGLKAVRSLVSATSLTKNGLALAGG-------YAGVS 356
Db    291 ----AILTGGEVITEDLGLDLKSANITQL-GRASKVVVTKENTTIVEGAGESDKIAARVN 345

Qy    357 KLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKST 416
Db    346 QIKAQIEETTSDFDKEKLQERLAKLAGGVAVL----KVGAATETEMKERKLRIEDALNST 401

Qy    417 ASSTTSYV----ADQTVKLAKTVKDMSGEAISSTGA-----SLRSTVNNLRHRSAPEADI 467
Db    402 RAAVEEGIVAGGGTALVNVIKAVSSIGAEGDEATGVNIVLRALEEPVRQIAHNAGLEGSV 461

Qy    468 -------EEGG 471
Db    462 IVERLKKEEAG 472
```

```
RESULT 9
HIPR_MOUSE
ID   HIPR_MOUSE     STANDARD;      PRT;  1068 AA.
AC   Q9JKY5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Huntingtin interacting protein 1 related (Hip1-related).
GN   HIP1R.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20082888; PubMed=10613908;
RA   Engqvist-Goldstein A.E.Y., Kessels M.M., Chopra V.S., Hayden M.R.,
RA   Drubin D.G.;
RT   "An actin-binding protein of the Sla2/Huntingtin interacting protein 1
RT   family is a novel component of clathrin-coated pits and vesicles.";
RL   J. Cell Biol. 147:1503-1518(1999).
CC   -!- FUNCTION: COMPONENT OF CLATHRIN-COATED PITS AND VESICLES, THAT MAY
CC       LINK THE ENDOCYTIC MACHINERY TO THE ACTIN CYTOSKELETON.
CC   -!- SUBUNIT: INTERACTS WITH ACTIN AND HUNTINGTIN INTERACTING PROTEIN 1
CC       (HIP1). DOES NOT INTERACT WITH HUNTINGTIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MEMBRANE-ASSOCIATED PROTEIN,
CC       MAINLY LOCALIZED AT THE ENDOCYTIC COMPARTMENTS AND IN THE
CC       PERINUCLEAR REGION.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. EXPRESSED AT LOWER LEVELS IN
CC       SKELETAL MUSCLE AND HEART. THE LEVEL OF EXPRESSION DOES NOT CHANGE
CC       APPRECIABLY DURING DEVELOPMENT.
CC   -!- DOMAIN: THE TALIN-LIKE DOMAIN BINDS F-ACTIN.
CC   -!- SIMILARITY: BELONGS TO THE SLA2 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF221713; AAF34662.1; -.
DR   MGD; MGI:1352504; Hip1r.
DR   InterPro; IPR001026; ENTH.
DR   InterPro; IPR002558; ILWEQ.
DR   Pfam; PF01417; ENTH; 1.
DR   Pfam; PF01608; I_LWEQ; 1.
DR   ProDom; PD011820; ILWEQ; 1.
DR   SMART; SM00273; ENTH; 1.
DR   SMART; SM00307; ILWEQ; 1.
KW   Actin-binding; Coiled coil.
FT   DOMAIN       77     80       POLY-SER.
FT   DOMAIN      311    316       POLY-GLU.
FT   DOMAIN      346    644       COILED COIL (POTENTIAL).
FT   DOMAIN      800   1068       TALIN-LIKE.
SQ   SEQUENCE   1068 AA;  119484 MW;  97CEE9D92CDF5DB1 CRC64;

  Query Match             5.0%;  Score 121.5;  DB 1;  Length 1068;
  Best Local Similarity   22.7%;  Pred. No. 5.7;
  Matches  119;  Conservative   74;  Mismatches  178;  Indels  153;  Gaps   29;

Qy      1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
Db    505 MKMEEQSDQLEKLKRELAARAGELARAQEAL-SRTEQSGSELSSRLDTLN-----AEKEA 558

Qy     61 VIGSPDERDAALAH-----NEQIDALVETRANRLYSEGETPATIAETFAKAEKF------ 109
Db    559 LSGVVRQREAELLAAQSLVREKEEALSQEQQRSSQEKGELRGQLAEKESQEQGLRQKLLD 618

Qy    110 DRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPL------TP--LISGALSG 161
Db    619 EQLAVLRS--------AAAEAEAILQDAVSKLD------DPLHLRCTSSPDYLVSRA-QA 663

Qy    162 AMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFS--- 218
Db    664 ALDSVS----GLEQGHTQYLASSEDA--SALVAALTRFSHLAADTIVN-GAATSHLAPTD 716

Qy    219 ---------------ALNVV-----RTVLAPALAS--RPSVQGAVDFGVSTAGGLVANAG 256
Db    717 PADRLMDTCRECGARALELVGQLQDQTVLRRAQPSLMRAPLQGILQLGQDL--------- 767

Qy    257 FGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKAIKSA------SYSGAA 310
Db    768 ---KPKSLDVRQEELG-----AMVDKE-MAATSAAIE--DAVRRIEDMMSQARHESSGVK 816

Qy    311 LNAGKRMAGLPLDVATDGLKAVRSLV-SATSLTKNGLALAGGYAGVSKLQKMATKNI--T 367
Db    817 LEVNERI----LNSCTDLMKAIRLLVMTSTSLQKE---IVESGRGAATQQEFYAKNSRWT 869

Qy    368 D---SATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVK---------S 415
Db    870 EGLISASKA---------------VGWGATQL-----VESADKVVLHMGKYEELIVCSHE 909

Qy    416 TASSTTSYVADQTVK------------LAKTVKDMSGEAISST 446
Db    910 IAASTAQLVAASKVKANKNSPHLSRLQECSRTVNERAANVVAST 953
```

```
RESULT 10
YM96_YEAST
ID   YM96_YEAST     STANDARD;      PRT;  1140 AA.
AC   Q04893;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 113.1 kDa protein in PRE5-FET4 intergenic region.
GN   YMR317W OR YM9924.09.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
```

```
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z54141; CAA90835.1; -.
DR   SGD; S0004936; YMR317W.
KW   Hypothetical protein; Repeat.
SQ   SEQUENCE   1140 AA;  113070 MW;  0153EBCA24FE5427 CRC64;

  Query Match             5.0%;  Score 120.5;  DB 1;  Length 1140;
  Best Local Similarity   19.8%;  Pred. No. 7.2;
  Matches  118;  Conservative   84;  Mismatches  233;  Indels  161;  Gaps   26;

Qy      12 GVAMESFRTASDAS--LASSSVRSV----STTSCRDLQAITDYLKHHVFAAHRFSVIGSP 65
           | |: |  | ||||  : |||   |    |:||      |:: |    |::   |:  ||
Db     203 GSAVTSVGTTSDASKEVFSSSTSDVSSLLSSTSSPASSTISETLP---FSSTILSITSSP 259

Qy      66 DERDAALAHNEQIDALVETRANRLYS----------EGETPATIA--------ETFAKA 106
             :|  | :  : :   :  :  |             | |:: :         : : :
Db     260 VSSEAPSATSSSVSSEASSSTSSSVSSEAPLATSSVVSSEAPSSTSSVVSSEAPSSTSSS 319

Qy     107 EKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLA------------TPLKPLTP 153
           :  :||:||     | | :||:    |:       :                ||   :
Db     320 VSSEISSTTSSSVSSEAPLATSSVVSSEAPSSTSSSVSSEISSTTSSSVSSEAPLATSSV 379

Qy     154 LISGALSGAMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVSV-KRHSPALGRQV----- 207
           : | | |     | ::       :   | :|    :::  :    |  :  :|
Db     380 VSSEAPSSTSSSVSSEAPSSTSSSVS--SEAPSSTSSSVSSEISSTKSSVMSSEVSSATS 437

Qy     208 ----------VDMGIAVQTFSALN--VVRTVLA-------------PALASRPSVQGAV 241
                     :    : : ||: |  |  |::|                |||   :: ::
Db     438 SLVSSEAPSAISSLASSRLFSSKNTSVTSTLVATEASSVTSSLRPSSETLASNSIIESSL 497

Qy     242 DFG----VSTAGGLVANAGFGDRMLSVQSR----------DQLRGGAFVLGMKD---KEP 284
             |    |||     |::  | :: |  ||             |   : : |        |
Db     498 STGYNSTVSTTTS-AASSTLGSKVSSSNSRMATSKTSSTSSDLSKSSVIFGNSSTVTTSP 556

Qy     285 KAALSEETD-----WLD-----AYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRS 334
           |::|         | |     |     : : | |  :    :|  | |       | |
Db     557 SASISLTASPLPSVWSDITSSEASSISSNLASSSAPSDNNSTIASASLIVTKTKNSVVSS 616

Qy     335 LV---------------SATSLTKNGLAL--------------AGGYAGVSKLQKMAT 363
           :|               |:|||  |                 | :: :|    : |
Db     617 IVSSITSSETTNESNLATSSTSLLSNKATARSLSTSNATSASNVPTGTFSSMSSHTSVIT 676

Qy     364 KNITDSATKAAVSQLSNLVGSVGVFAGW---------TTAGLATDPAVKKAES------F 408
             : |:   |::  | :| |    ||:          ||: | |  |      |       |
Db     677 PGFSTSSASLAIN--STVVSS--SLAGYSFSTPESSPTTSTLVTSEAPSTVSSMTTSAPF 732

Qy     409 IQDKVKSTAS-STTSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAP 463
           | :   :  | || |:: : |  :: :|   ||:  ||  |      :||   |||
Db     733 INNSTSARPSPSTASFITESTSSIS-SVPLASGDVTSSLAA------HNLTTFSAP 781


RESULT 11
MSB2_YEAST
ID   MSB2_YEAST     STANDARD;      PRT;  1306 AA.
AC   P32334;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MSB2 protein (Multicopy suppression of a budding defect 2).
GN   MSB2 OR YGR014W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92383951; PubMed=1514328;
RA   Bender A., Pringle J.R.;
RT   "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
RT   defect.";
RL   Yeast 8:315-323(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=97435481; PubMed=9290212;
RA   Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT   "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT   chromosome VII.";
RL   Yeast 13:1077-1090(1997).
CC   -!- PTM: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).
CC   -!- SIMILARITY: SOME, TO YEAST HKR1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M77354; AAA34798.1; -.
DR   EMBL; Z72799; CAA96997.1; -.
DR   PIR; S25370; S25370.
DR   SGD; S0003246; MSB2.
KW   Transmembrane; Glycoprotein; Repeat.
FT   DOMAIN      698    816       7 X 17 AA TANDEM REPEATS.
FT   REPEAT      698    714       1.
FT   REPEAT      715    731       2.
FT   REPEAT      732    748       3.
FT   REPEAT      749    765       4.
FT   REPEAT      766    782       5.
FT   REPEAT      783    799       6.
FT   REPEAT      800    816       7.
SQ   SEQUENCE   1306 AA;  133114 MW;  67D5D984D5CA4A6D CRC64;

  Query Match             5.0%;  Score 120;  DB 1;  Length 1306;
  Best Local Similarity   20.2%;  Pred. No. 9.1;
  Matches  117;  Conservative   79;  Mismatches  235;  Indels  148;  Gaps   24;

Qy      14 AMESFRTASDASLAS----SSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERD 69
           |  |    ||| | ||    :|  |:  |:  :::  |                 :|   |
Db     142 AASSVNAASDVSTASVPIDTSANSIPFTTTSNIETTT---------------SAPLTSD 185

Qy      70 AALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENT-PFAAA 128
             |     : |     |: ::|     :    |   :| ||: :|  :   :::  |:::
Db     186 TPLISTSTMSA-----ADNVFSSANPISASLTTTDSSESFDQTSTAGAIPVQSSADFSSS 240

Qy     129 SVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVG-TKMMDRARGDLHYLSTSPDK 187
           | :     |      ::|  | |  ||  || |  |   :     |   |   ||  |
Db     241 SEILVQSSAD-----FSSPSSPTTTDIS--LSAAPLQTSESSSFTTASAALPVSSTDVDG 293

Qy     188 LHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNV-------------------VRT--- 225
             :  ||:              |     :|||  :                    ::|
Db     294 SSASPVVSMSAAGQIASSSSTDNPTMSETFSLTSTEVDGSDVSSTVSALLSAPFLQTSTS 353

Qy     226 ----VLAPALASRPSVQGAVDFGVSTAGGLVANAGFGD-----------------RMLS 263
                :::|:::  || | : |   |:   :|::: | |                  |:
Db     354 NSFSIVSPSVSFVPS-QSSSDVASSSTANVVSSS-FSDIPPQTSTSGSVVSVAQSASALA 411

Qy     264 VQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAG-KRMAGLPL 322
           ||  :: | :    |       :|   |  ||:     |:: |    : |    | :||
```

```
Db    412 FQSSTEVYGASASSTMSSLLSTTSLQSTT--LDSSSLASSSASSSDLTDYGVSSTASIPL 469

Qy    323 DVATDGLKAVRSL--------------------VSATSLTKNGLALAGGYAGVSKLQKMA 362
Db    470 LSASEQASTSSSFSVVSPSVSFVPSQSSSDVASTSAPSVVSSSFSYTSLQAGGSSMTNPS 529

Qy    363 TKNI-----TDSATKAAVSQLS-NLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKST 416
Db    530 SSTIVYSSSTGSSEESAASTASATLSGSSSTY---MAGNLQSQP--PSTSSLLSE---SQ 581

Qy    417 ASSTTSYVADQTVKLAK---TVKDMSGEA---ISSTGA-----SLRSTVNNLRHRSAPEA 465
Db    582 ATSTSAVLASSSVSTTSPYTTAGGASTEASSLISSTSAETSQVSYSQSTTALQTSSFASS 641

Qy    466 DIEEGG--------------------ISAFSRSETPFQL 484
Db    642 STTEGSETSSQGFSTSSVLVQMPSSISSEFSPSQTTTQM 680


RESULT 12
HLYA_SERMA
ID   HLYA_SERMA     STANDARD;      PRT;  1608 AA.
AC   P15320;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-NOV-1990 (Rel. 16, Last annotation update)
DE   Hemolysin precursor.
GN   SHLA.
OS   Serratia marcescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Serratia.
OX   NCBI_TaxID=615;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RC   STRAIN=SN8;
RX   MEDLINE=88257037; PubMed=3290200;
RA   Poole K., Schiebel E., Braun V.;
RT   "Molecular characterization of the hemolysin determinant of Serratia
RT   marcescens.";
RL   J. Bacteriol. 170:3177-3188(1988).
CC   -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC       CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC       DEFINED.
CC   -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC       ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
CC       REQUIRES SHLB FUNCTION.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).
CC   ---------------------------------------------------------------------
CC . This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M22618; AAA50323.1; -.
DR   PIR; A28182; A28182.
KW   Hemolysis; Toxin; Outer membrane; Signal.
FT   SIGNAL       1     30
FT   CHAIN        31   1608       HEMOLYSIN.
SQ   SEQUENCE   1608 AA;  165078 MW;  D669B476FE7DAD51 CRC64;

  Query Match             5.0%;  Score 119.5;  DB 1;  Length 1608;
  Best Local Similarity   20.5%;  Pred. No. 13;
  Matches  120;  Conservative   68;  Mismatches  242;  Indels  155;  Gaps   21;

Qy      8 QQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDE 67
Db    708 QYRAGLRIEHTRDSEKTTRTENSASSLSGGSVK-LKAEKDV----TFSGSKLVA----DK 758

Qy     68 RDAALAHNEQIDALVETRANRLYSEGETPATIAET-----FAKAEKFDRLATTASSAFEN 122
Db    759 GDASVSGN---------KVSFLAADDKTASNTEQTKIGGGFYYTGGIDKLGSGVEAGYEN 809

Qy    123 TPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY-- 180
Db    810 NKTQAQSSKAITSGSDVKGNLTINARDKLTQ------QGAQHSVGGAYQENAAGVDHLAA 863

Qy    181 ---LSTSPDKLHDAMAVSVKRHSPALGRQV---VDMGIAVQTFSALNVVRTVLAPALASR 234
Db    864 ADTASTTTTKTDVGVNIGANVDYSAVTRPVERAVGKAAKLDATGVINDIGGIGAPNVGLD 923

Qy    235 PSVQGAVDFGVSTAGGLVAN---AGFGDRMLSVQSRDQ---------------------- 269
Db    924 IGAQGGSSEKRSSSSQAVVSSVQAGSIDINAKGEVRDQGTQYQASKGAVNLTADSHRSEA 983

Qy    270 -----------LRGGAFVL----------------GMKDKEPKAALSEETDWLDAYKAI- 301
Db    984 AANRQDEQSRDTRGSAGVRVYTTTGSDLTVDAKGEGGTQRSNSSASQAVTGSIDAANGIN 1043

Qy    302 ----KSASYSGAALNAGKRM----AG--LPLDVATDGLKAVRSLVSATSLTKNGLA---- 347
Db   1044 VNVKKDAIYQGTALNGGRGKTAVNAGGDIRLDQASDKQSESRSGFNVKASAKGGFTADSK 1103

Qy    348 -LAGGYAGVSKLQKMATKNITDSATKAAVSQLS--------------------NLVGSVG 386
Db   1104 NFGAGFGG-------GTHNGESSSSTAQVGNISGQQGVELKAGRDLTLQGTDVKSQGDVS 1156

Qy    387 VFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAISST 446
Db   1157 LSAGNKVALQAAESTQTRKESKLSGNIDLGAGSSDS------------KEKTGGNLSAG 1203

Qy    447 GASLRSTVNNLRHRSAPE---ADIEEGG---ISAFSRSETPFQLR 485
Db   1204 GAFDIAKVN----ESATERQGATIASDGKVTLSANGKGDDALHLQ 1244


RESULT 13
HS7F_CAEEL
ID   HS7F_CAEEL     STANDARD;      PRT;   657 AA.
AC   P11141; P91135;
DT   01-JUL-1989 (Rel. 11, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Heat shock 70 kDa protein F, mitochondrial precursor.
GN   HSP-6 OR HSP70F OR C37H5.8.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Davidson S., Gillam B.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-487 FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=89356251; PubMed=2766926;
RA   Heschl M.F.P., Baillie D.L.;
RT   "Characterization of the hsp70 multigene family of Caenorhabditis
RT   elegans.";
RL   DNA 8:233-243(1989).
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U88315; AAB42371.1; -.
DR   EMBL; X07678; CAA30525.1; -.
DR   PIR; S03210; S03210.
DR   PIR; B32475; B32475.
DR   HSSP; P04475; 1DG4.
DR   WormPep; C37H5.8; CE08631.
DR   InterPro; IPR001023; Hsp70.
DR   Pfam; PF00012; HSP70; 1.
DR   PRINTS; PR00301; HEATSHOCK70.
DR   ProDom; PD000089; Hsp70; 1.
DR   PROSITE; PS00297; HSP70_1; 1.
DR   PROSITE; PS00329; HSP70_2; 1.
DR   PROSITE; PS01036; HSP70_3; 1.
KW   ATP-binding; Heat shock; Multigene family; Mitochondrion;
KW   Transit peptide.
FT   TRANSIT       1     27       MITOCHONDRION (POTENTIAL).
FT   CHAIN        28    657       HEAT SHOCK 70 KDA PROTEIN F.
FT   CONFLICT    138    138       V -> E (IN REF. 2).
FT   CONFLICT    140    140       S -> P (IN REF. 2).
FT   CONFLICT    167    167       A -> P (IN REF. 2).
FT   CONFLICT    347    348       KA -> NV (IN REF. 2).
SQ   SEQUENCE   657 AA;  70844 MW;  637A1636B6ABB836 CRC64;

  Query Match             4.9%;  Score 119;  DB 1;  Length 657;
  Best Local Similarity  21.5%;  Pred. No. 4.3;
  Matches  107;  Conservative   60;  Mismatches  162;  Indels  168;  Gaps   26;

Qy      29 SSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDALVETRANR 88
           || |::: :|    ::::|  | ||                |       :::| :  :
Db       9 SSARTIARSSLMSARSLSDKPKGHVIGID-------------LGTTNSCVSIMEGKTPK 54

Qy      89 LYSEGE----TPATIAETFAKAEKF-----DRLATTASSAFENTPFAAASVL--QYMQPA 137
           :    |    ||:|:| | |  |:        | | | |:   || ||   ::  :|  |
Db      55 VIENAEGVRTTPSTVAFT-ADGERLVGAPAKRQAVTNSA---NTLFATKRLIGRRYEDPE 110

Qy     138 INK--------------GD-WLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLS 182
           : |              || |:    |  :|   ||       |  || : |     ||
Db     111 VQKDLKVVPYKIVKASNGDAWVEAQGKVYSPSQVGAF------VLMKMKETAES---YLG 161

Qy     183 TSPDKLHDAMAVSVKRHSPALGRQVV-DMGIAVQTFSALNVVRTVLAPALASRPSVQGAV 241
           |: :      |:|  :      ||   | |      | |||:| :  |  |
Db     162 TTVNN----AVVTVPAYFNDSQRQATKDAG----QISGLNVLRVINEPTAA--------- 204

Qy     242 DFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKAI 301
               : | ||  :|  ||:::::|     | || |               :   |:  |
Db     205 ----ALAYGLDKDA--GDKIIAVY---DLGGGTF--------------DVSILEIQK-- 238

Qy     302 KSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKM 361
                                   |:  |:|         ||    ||
Db     239 -------------------------GVPEVKS--------TNGDTFLGG---------- 254

Qy     362 ATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTT 421
               :   | | |      || |       | ||   :      :| |   ||||
Db     255 ------EDFDHALVHHL------VGEFKKEQGVDLTKDPQAMQRLREAAEKAKCELSSTT 302

Qy     422 S------YVA-DQT------VKLAKT-VKDMSGEAISSTGASLRSTVNNLRHRSAPEADI 467
               |:  ||:      :|| :   : : |: | |    |  :::   :|:  ||:
Db     303 QTDINLPYITMDQSGPKHLNLKLTRAKFEQIVGDLIKRTIEPCRKALHDAEVKSSQIADV 362

Qy     468 -EEGGISAFSRSETPFQ 483
            ||:|   : :  |
Db     363 LLVGGMSRMPKVQATVQ 379


RESULT 14
METE_MYCLE
ID   METE_MYCLE     STANDARD;      PRT;   760 AA.
AC   O05564;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
DE   (EC 2.1.1.14) (Methionine synthase, vitamin-B12 independent isozyme)
DE   (Cobalamin-independent methionine synthase).
GN   METE OR ML0961 OR MLCB33.14.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM 5-
CC       METHYLTETRAHYDROFOLATE TO HOMOCYSTEINE RESULTING IN METHIONINE
CC       FORMATION (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 5-methyltetrahydropteroyltri-L-glutamate + L-
CC       homocysteine = tetrahydropteroyltri-L-glutamate + L-methionine.
CC   -!- COFACTOR: ZINC; BINDS ONE ION PER SUBUNIT (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.
CC   -!- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE
CC       SYNTHASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z94723; CAB08123.1; -.
DR   EMBL; AL583920; CAC31342.1; -.
DR   Leproma; ML0961; -.
DR   InterPro; IPR002629; Methionine_synt.
DR   Pfam; PF01717; Methionine_synt; 1.
KW   Transferase; Methyltransferase; Methionine biosynthesis; Zinc; Repeat;
KW   Complete proteome.
FT   METAL       647    647       ZINC (BY SIMILARITY).
FT   METAL       649    649       ZINC (BY SIMILARITY).
FT   METAL       732    732       ZINC (BY SIMILARITY).
SQ   SEQUENCE   760 AA;  82234 MW;  1E92FA570B9ACC97 CRC64;

  Query Match             4.9%;  Score 118.5;  DB 1;  Length 760;
  Best Local Similarity  21.1%;  Pred. No. 5.6;
  Matches   91;  Conservative   66;  Mismatches  173;  Indels  101;  Gaps   18;

Qy      85 RANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWL 144
           ||   |    |  :  |: |   : |  :  |::  :: |    :   |    ::
Db      28 RATEGYWAKRTSRSELESVASTLRRDMWSDLAAAGLDSVP---VNTFSYYDQMLD----- 79

Qy     145 ATPLKPLTPLISGALSGAMDQVGTKM-----MDRARGDL-------------HYLSTSPD 186
                 |  : |||   : ||   :     : |   |:              |||
Db      80 -------TAFMLGALPARVAQVSDDLDQYFALARGNNDIKPLEMTKWFDTNYHYL----- 127

Qy     187 KLHDAMAVSVKRHSPALGRQVVDMGIAVQTFS-----ALNVVRTVLAPALASRPSVQGAV 241
                 |    ||              |||      |  ::  |   : ||| : | |
Db     128 ---------VPEIEPA-----------TTFSLNPGKILGELKEALEQRIPSRPVIIGPV 166

Qy     242 DF-----GVSTAGGLVAN----AGFGDRMLSVQSRD-----QLRGGAFVLGMKDKEPKAA 287
           |     |::  |  :        |    :||: : :     |    | |  :    | |
Db     167 TFLLLSKGINGGGAPIQRLEELVGIYCTLLSLLAENGARWVQFDEPALVTDLSPDAPALA 226

Qy     288 LSEETDWLDAYK--AIKSASYSGAALNAGKRMAGL---PLD-VATDGLKAVRSLVSATSL 341
             :  |      |  ||  |:| |   | |  :|||   |:: :  | :    : |:|
```

```
Db    227 EAVYTALGSVSKRPAIYVATYFG---NPGASLAGLARTPIEAIGVDFVCGADTSVAAVPE 283

Qy    342 TKNGLALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPA 401

Db    284 LAGKTLVAGIVDG---------RNIWRTDLESALSKLATLLGSAATVAVSTSCSTLHVPY 334

Qy    402 VKKAESFIQDKVKSTASSTTSYVADQTVKLAKTV---KDMSGEAISSTGASLRSTVNNLR 458

Db    335 SLEPETDLDDNLRSWLAFGAEKVA-EVVVLARALRDGRDAVADEIAASNAAVASRRSDPR 393

Qy    459 -HRSAPEADIE 468

Db    394 LHNGQVRARID 404


RESULT 15
LYTD_BACSU
ID   LYTD_BACSU     STANDARD;      PRT;   880 AA.
AC   P39848;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Beta-N-acetylglucosaminidase precursor (EC 3.2.1.96).
GN   LYTD OR CWLG.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95020588; PubMed=7934877;
RA   Margot P., Maueel C., Karamata D.;
RT   "The gene of the N-acetylglucosaminidase, a Bacillus subtilis 168
RT   cell wall hydrolase not involved in vegetative cell autolysis.";
RL   Mol. Microbiol. 12:535-545(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AC327;
RX   MEDLINE=96036197; PubMed=7581999;
RA   Rashid M.H., Mori M., Sekiguchi J.;
RT   "Glucosaminidase of Bacillus subtilis: cloning, regulation, primary
RT   structure and biochemical characterization.";
RL   Microbiology 141:2391-2404(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: CELL WALL HYDROLASE NOT INVOLVED IN CELL AUTOLYSIS. IT
CC       HYDROLYZES THE BETA-1,4 GLYCAN BOND BETWEEN THE N-
CC       ACETYLGLUCOSAMINYL AND THE N-ACETYLMURAMOYL RESIDUES IN THE GLYCAN
CC       CHAIN.
CC   -!- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC       unit in high-mannose glycopeptides and glycoproteins containing
CC       the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC       residue remains attached to the protein; the rest of the
CC       oligosaccharide is released intact.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U02562; AAA67857.1; -.
DR   EMBL; D45048; BAA08089.1; -.
DR   EMBL; Z99122; CAB15595.1; -.
DR   SubtiList; BG10455; lytD.
DR   InterPro; IPR002901; Amidase_4.
DR   InterPro; IPR003646; SH3_bac.
DR   Pfam; PF01832; Amidase_4; 1.
DR   SMART; SM00047; LYZ2; 1.
DR   SMART; SM00287; SH3b; 1.
KW   Cell wall; Hydrolase; Signal; Repeat; Complete proteome.
FT   SIGNAL       1     27       POTENTIAL.
FT   CHAIN       28    880       BETA-N-ACETYLGLUCOSAMINIDASE.
FT   DOMAIN      72     75       POLY-THR.
FT   DOMAIN     337    340       POLY-LYS.
FT   DOMAIN     568    571       POLY-ALA.
FT   REPEAT     439    473       1.
FT   REPEAT     479    513       2.
SQ   SEQUENCE   880 AA;  95553 MW;  2A912A478FCFC1D1 CRC64;

 Query Match             4.9%;  Score 118.5;  DB 1;  Length 880;
 Best Local Similarity   18.3%;  Pred. No. 6.8;
 Matches   90;  Conservative   77;  Mismatches  185;  Indels  141;  Gaps   15;

Qy     26 LASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDA----- 80

Db     11 LSAASLAFFAMSGSAQAAAYTDYSLYKVEPSNTFSTESQASQAVAKLEKDTGWDASYQAS 70

Qy     81 ----LVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQP 136

Db     71 GTTTTYQISASGIHSESEAKAILS----------GLAKQTSITGTSSPVGSK-------- 112

Qy    137 AINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVSV 196

Db    113 ------------QPYVTISSGAISG----------------------EKQANTILAKL 136

Qy    197 KRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVANAG 256

Db    137 KQETGVAG--------AVKAYGAAQPYMNVMTSDIADETKVKALIQ-SLAKQTGIKSSYQ 187

Qy    257 FGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWL---------DAYKAIKSASYS 307

Db    188 PITHTVSVTT---IQSGTIVGDSRAAQIKNAFQKESGLQASLKETVKGQAYYTFTTAAIS 244

Qy    308 GAA-----LNAGKRMAGL-------------PLDVATDGLKAVRSLVSATSLTKNGLAL 348

Db    245 GEANAKTLLQQLKQSTGITGSYKSINQKTTVESYNVQSAYFKGLSTVKDAISQIKKNTGV 304
```

```
Qy     349 AGGYAGVSK---------------LQKMAT------KNITDSATKAAVSQLSNLVGSVGV 387
           :| |  | |               |||: |      : | |: |   :  :  :
Db     305 SGSYQQVGKSTSYTVNMKGITKQQLQKIDTFFKKKKWHYTSSSVKKTTTSAAYQI----- 359

Qy     388 FAGWTTAGLATDPAVKKAESFI-QDKVK-------STASSTTSYVADQTVKLAKTVKDMS 439
             ||| :  :    || :|  | |||       ||| :    ::::|    ||  | ::
Db     360 ----TTAKILGEQQANKAAAFFAQKKVKAAKTAAGSTAENQYQLISEETSDQAKVTKGLN 415

Qy     440 GEAISSTGASLRS 452
             :   || :|
Db     416 ILKKNQLSASAKS 428
```

Search completed: January 2, 2003, 15:19:24
Job time : 14.5098 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:17:10 ; Search time 17.0175 Seconds
(without alignments)
2751.141 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINQSAQQPPGVAMESFRT.....EEGGISAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	149.5	6.2	2155	2	C97523	hypothetical prote
2	149.5	6.2	2155	2	AD2742	conserved hypothet
3	149	6.2	989	2	AE3045	ice nucleation pro
4	149	6.2	1009	2	G98240	hypothetical prote
5	143.5	6.0	2055	2	T31110	extracellular matr
6	141	5.9	536	2	D84325	Htr17 transducer [
7	140.5	5.8	2186	2	H89960	hypothetical prote
8	140	5.8	2285	2	T12796	probable transglyc
9	138	5.7	1122	2	G64887	probable tail fibe
10	138	5.7	1731	2	B98241	hypothetical prote
11	138	5.7	1731	2	AB3045	ice nucleation pro
12	135	5.6	971	2	B90835	probable tail fibe
13	135	5.6	973	2	C85693	probable membrane
14	134	5.6	2535	2	AC0304	probable hemolysin
15	132.5	5.5	654	2	C87587	hypothetical prote
16	132.5	5.5	1104	2	S59310	probable membrane
17	131.5	5.5	1214	2	G97419	streptococcal hema
18	131.5	5.5	1248	2	AH2637	conserved hypothet
19	131	5.4	409	2	S70548	sspC protein - Sal
20	130	5.4	1026	2	A48995	paracrystalline su
21	130	5.4	1073	2	C87374	S-layer protein Rs
22	129	5.4	436	2	G81655	conserved hypothet
23	129	5.4	637	2	C87322	methyl-accepting c
24	128.5	5.3	2388	2	JE0271	beta spectrin, bet
25	127	5.3	397	2	I39579	nccB protein - Alc
26	127	5.3	1010	2	AH2553	hypothetical prote
27	126.5	5.3	1131	2	T41144	hypothetical serin
28	126.5	5.3	1630	2	A53577	ascites sialoglyco
29	126.5	5.3	2271	2	F90073	hypothetical prote
30	125.5	5.2	902	2	H87323	hypothetical prote
31	125.5	5.2	1225	2	F75605	minor tail protein
32	125	5.2	409	2	S70215	sipC protein - Sal
33	125	5.2	409	2	AH0850	pathogenicity isla
34	125	5.2	1329	2	T29074	hypothetical prote
35	124	5.2	1128	1	T08322	plasmid replicatio
36	124	5.2	2481	2	D90011	FmtB protein [impo
37	123.5	5.1	860	2	T14650	tail fiber protein
38	123.5	5.1	978	2	T14968	phage lambda-relat
39	123.5	5.1	1147	2	T35781	hypothetical prote
40	123.5	5.1	2232	2	T34434	hypothetical prote
41	123	5.1	544	2	JC6063	chaperonin groEL -
42	123	5.1	544	2	B83720	class I heat-shock
43	122.5	5.1	789	2	E84236	Htr6 transducer [i
44	122	5.1	6713	2	B89921	hypothetical prote
45	121.5	5.0	518	2	A55864	flagellin - Woline

ALIGNMENTS

RESULT 1
C97523
hypothetical protein AGR_C_2490 [imported] - Agrobacterium tumefaciens (strain C58, Cer
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: C97523
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B., Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
A;Reference number: A97359; PMID:11743194
A;Accession: C97523
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2155 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87140.1; PID:g15156408; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_2490
A;Map position: circular chromosome

Query Match 6.2%; Score 149.5; DB 2; Length 2155;
Best Local Similarity 20.7%; Pred. No. 0.35;
Matches 96; Conservative 65; Mismatches 176; Indels 127; Gaps 16;

```
Qy      69 DAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF----ENTP 124
           ||    : :|| ::  |:| |:             |  ::||    : | |       |
Db     609 DAFTNSHAKIDTVLAERSNALF---------GALSASQDRFDEALASRSLAITGSVSGTA 659

Qy     125 FAAASVLQYMQPAINK-----GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLH 179
             |::|     |||          |   |:     |:||:||  |::   :  |
Db     660 EHLAAMLDERAAAINSVVADVERRLTETLETRAAAITGAVSGIEDRISDTLESRTAA--- 716

Qy     180 YLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASR----- 234
               ||| :: :  | :  |       |       ||:: |   :|  : ||
Db     717 --------LHDVVSGAESRIADTLD------GRTAALSSAISGVEERIADTMDSRTLSLD 762

Qy     235 -------PSVQGAVDFGVSTAGGLVAN-------------AGFGDRMLSVQSR--DQLRG 272
                  :   :|   |   |:||:             | ||| :   ::|  : | |
Db     763 MTFANVEERLSETLDNRTSALTGIVASAEEKIAGALDSRTATFGDVVAGAETRIAETLDG 822

Qy     273 -----GAFVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGA----------------- 309
               | | | :::          | ||:       ::|||
Db     823 RTAALNAVVSGAEER--------IADALDSRTMALDMTFSGAEEKIAEALDTRTAALGEL 874

Qy     310 ALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDS 369
             :|  |:||  ||  || || |                       ||  ::  | :: ||
Db     875 VASAETRIAG-ALDSRTDSLKTV---------------------VSGAEERIT-DVLDS 910

Qy     370 ATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTV 429
             | |     | :   :       ||       |:| | : ::|::     | |: ::
Db     911 RTMALDMSFSGVEEKITDILDGRTA------ALKSAVAGVEDRIAGALDSRTAALSG--- 961
```

```
Qy    430 KLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGIS 473
Db    962 -IVSGAEERIAEALDSRTLALDMTISGVEERIAEAMDARASSLS 1004
```

RESULT 2
AD2742
conserved hypothetical protein Atu1348 [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AD2742
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AD2742
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2155 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL42354.1; PID:g17739760; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu1348
A;Map position: circular chromosome

Query Match 6.2%; Score 149.5; DB 2; Length 2155;
Best Local Similarity 20.7%; Pred. No. 0.35;
Matches 96; Conservative 65; Mismatches 176; Indels 127; Gaps 16;

```
Qy     69 DAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAF----ENTP 124
Db    609 DAFTNSHAKIDTVLAERSNALF---------GALSASQDRFDEALASRSLAITGSVSGTA 659

Qy    125 FAAASVLQYMQPAINK-----GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLH 179
Db    660 EHLAAMLDERAAAINSVVADVERRLTETLETRAAAITGAVSGIEDRISDTLESRTAA--- 716

Qy    180 YLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASR----- 234
Db    717 --------LHDVVSGAESRIADTLD------GRTAALSSAISGVEERIADTMDSRTLSLD 762

Qy    235 -------PSVQGAVDFGVSTAGGLVAN-------------AGFGDRMLSVQSR--DQLRG 272
Db    763 MTFANVEERLSETLDNRTSALTGIVASAEEKIAGALDSRTATFGDVVAGAETRIAETLDG 822

Qy    273 -----GAFVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGA----------------- 309
Db    823 RTAALNAVVSGAEER--------IADALDSRTMALDMTFSGAEEKIAEALDTRTAALGEL 874

Qy    310 ALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDS 369
Db    875 VASAETRIAG-ALDSRTDSLKTV----------------------VSGAEERIT-DVLDS 910

Qy    370 ATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTV 429
Db    911 RTMALDMSFSGVEEKITDILDGRTA------ALKSAVAGVEDRIAGALDSRTAALSG--- 961

Qy    430 KLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGIS 473
Db    962 -IVSGAEERIAEALDSRTLALDMTISGVEERIAEAMDARASSLS 1004
```

RESULT 3
AE3045
ice nucleation protein homolog [imported] - Agrobacterium tumefaciens (strain C58, Dupon
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AE3045
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AE3045
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-989 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44779.1; PID:g17742417; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: ina
A;Map position: linear chromosome

Query Match 6.2%; Score 149; DB 2; Length 989;
Best Local Similarity 22.4%; Pred. No. 0.13;
Matches 127; Conservative 77; Mismatches 228; Indels 136; Gaps 25;

```
Qy      7 AQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAI-TDYLKH---HVFAAHRFSVI 62
Db     98 ASQIDAMSTDQIKALNSSQVAGLSSAQVATLSSDELALFTTDEIKSISANAIAGLSAAAI 157

Qy     63 GSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKF------DRLATTA 116
Db    158 AGLSTDNAAALTKSQIAAMSSTQFNALTSGSLATFSADEVKAISNKILAGLDVTKLSTGN 217

Qy    117 SSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARG 176
Db    218 IAALSKAQVSALSTTQFAAMSTDQ-------IKALTSDQVAGLSSA--QVATLSSD---- 264

Qy    177 DLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIA---------------------VQ 215
Db    265 ELALFSTDEIKAIGANAVA-GLSAAALAALTTDNAAALTKTQIAGLSSTQLNALTSANLA 323

Qy    216 TFSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVA-----NAGFGDRMLSVQSRDQL 270
Db    324 TFSA-DEIKAISTKALA-------GLDVTKLSTGNIAALTKTQAASLSSTQFAAMSTDQI 375

Qy    271 RG--GAFVLGMKDKEPKAALSEE-----TDWLDAYKAIKSASYSGAALNA---------- 313
Db    376 KALTSEQVAGLSSAQVATLSSDELALFSTDEISAFSANAVAGLSTAALAALTTGNATALT 435

Qy    314 GKRMAGL---PLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSA 370
Db    436 KTQIAGLSSTQLNALTSGSIATFSADEVKAISTK--ALAG-----LDVTKLSTGNVA-AL 487

Qy    371 TKAAVSQLSNLVGSVGVFAGWTT-----------AGLATDPAVKKAES----FIQDKVKS 415
Db    488 SKAQVSALSTT-----QFAAMSTDQIKALTSEQVAGLSSAQVATLSSDELALFSTDEIKA 542

Qy    416 ---------TASSTTSYVADQTVKLAKT-VKDMSGEAISSTGASLRSTVNNLRHRSAPE- 464
Db    543 IGANAVAGLSAAALAALTTDNASALTKTQIAGMSSTQINAL------TSANLATFSADEI 596

Qy    465 -------------ADIEEGGISAFSRSE 479
Db    597 KAITTKALGGLDVTKLSTGNIAALTKAQ 624
```

RESULT 4
G98240
hypothetical protein AGR_L_1758 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: G98240
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194

A;Accession: G98240
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1009 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89449.1; PID:g15159312; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1758
A;Map position: linear chromosome

Query Match 6.2%; Score 149; DB 2; Length 1009;
Best Local Similarity 22.4%; Pred. No. 0.13;
Matches 127; Conservative 77; Mismatches 228; Indels 136; Gaps 25;

```
Qy       7 AQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAI-TDYLKH---HVFAAHRFSVI 62
Db     118 ASQIDAMSTDQIKALNSSQVAGLSSAQVATLSSDELALFTTDEIKSISANAIAGLSAAAI 177

Qy      63 GSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKF------DRLATTA 116
Db     178 AGLSTDNAAALTKSQIAAMSSTQFNALTSGSLATFSADEVKAISNKILAGLDVTKLSTGN 237

Qy     117 SSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARG 176
Db     238 IAALSKAQVSALSTTQFAAMSTDQ-------IKALTSDQVAGLSSA--QVATLSSD---- 284

Qy     177 DLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIA--------------------VQ 215
Db     285 ELALFSTDEIKAIGANAVA-GLSAAALAALTTDNAAALTKTQIAGLSSTQLNALTSANLA 343

Qy     216 TFSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVA-----NAGFGDRMLSVQSRDQL 270
Db     344 TFSA-DEIKAISTKALA-------GLDVTKLSTGNIAALTKTQAASLSSTQFAAMSTDQI 395

Qy     271 RG--GAFVLGMKDKEPKAALSEE-----TDWLDAYKAIKSASYSGAALNA---------- 313
Db     396 KALTSEQVAGLSSAQVATLSSDELALFSTDEISAFSANAVAGLSTAALAALTTGNATALT 455

Qy     314 GKRMAGL---PLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSA 370
Db     456 KTQIAGLSSTQLNALTSGSIATFSADEVKAISTK--ALAG-----LDVTKLSTGNVA-AL 507

Qy     371 TKAAVSQLSNLVGSVGVFAGWTT-----------AGLATDPAVKKAES----FIQDKVKS 415
Db     508 SKAQVSALSTT-----QFAAMSTDQIKALTSEQVAGLSSAQVATLSSDELALFSTDEIKA 562

Qy     416 ---------TASSTTSYVADQTVKLAKT-VKDMSGEAISSTGASLRSTVNNLRHRSAPE- 464
Db     563 IGANAVAGLSAAALAALTTDNASALTKTQIAGMSSTQINAL------TSANLATFSADEI 616

Qy     465 -------------ADIEEGGISAFSRSE 479
Db     617 KAITTKALGGLDVTKLSTGNIAALTKAQ 644
```

RESULT 5
T31110
extracellular matrix binding protein - Abiotrophia defectiva (fragment)
C;Species: Abiotrophia defectiva
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C;Accession: T31110
R;Manganelli, R.; van de Rijn, I.
Infect. Immun. 67, 50-56, 1999
A;Title: Cloning and characterization of emb, a gene encoding the major adhesin of Strep
A;Reference number: Z20988; MUID:99081722; PMID:9864195
A;Accession: T31110
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2055 <MAN>
A;Cross-references: EMBL:AF067776; NID:g3249002; PID:g3249003; PIDN:AAD03320.1
C;Genetics:
A;Gene: emb

Query Match 6.0%; Score 143.5; DB 2; Length 2055;
Best Local Similarity 22.2%; Pred. No. 0.79;
Matches 123; Conservative 78; Mismatches 249; Indels 105; Gaps 23;

```
Qy       3 INQSAQQP------PGVAMESFRTASDASLA--SSSVRSVSTTSCRD--LQAI----TDY 48
Db     726 INEISQRPDLTREEKQAFMDQVRTARDAAMAKVASAANNQAVTSARDQGLNAVNNLPTPA 785

Qy      49 LKHHVFAAHRFSVIGSPDERDAALAHN------EQIDAL-----VETRANRLYSEGETPA 97
Db     786 AKYPEALGH---VRQAADAKRQAIRDNANLTAEEQADALRQVDAAQTAAEAAINQNHTNA 842

Qy      98 TIAETFAKAEKF-------DRLATTASSAFENTPFAAASVLQYMQPAINKGDWL-----A 145
Db     843 TLAKADSDGVKAINDINPQPRSKPAANQALEQV--AAAK-----RQAINNNNQLTDEEKA 895

Qy     146 TPLKPLTPLISGA---LSGAMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPA 202
Db     896 QAIQQVDQALANAKTQVQAANDNNGVNQAKTA-GTTAINNINPQGTQKAQAIAAIEAAEQ 954

Qy     203 LGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQ---------GAVDFGVSTAGGLVA 253
Db     955 AKRLELQGRNDLTTEERNNALADLTAKAQAAKDAVNQARNNTGVAGAKDNGVAQIQGINP 1014

Qy     254 NA------------GFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWL-DAYKA 300
Db    1015 TAVVKPDARNAIDQAARDKEAEFQANTKL---------TDEEKAAAIKKVQDAARDAKAA 1065

Qy     301 IKSASYSG---AALNAGKRMAGLPLDVATDGLKAV-RSLVSATSLTKNGL---------A 347
Db    1066 IDRAGSNGDVNNAVNQGK--------AAIQAIKALDDSQPSAKDTAKAAIQNAADAKKAA 1117

Qy     348 LAGGYAGVSKLQKMATKNITDSATKA-AVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAE 406
Db    1118 ITANNALTQEEKAAAIKQVEDEAAKAQAAVDASRSKADVDRAKDQGLQKISDVPAVQPPK 1177

Qy     407 SFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEAD 466
Db    1178 LNAIAAVDQAATDKKAVINNDTT-LTQEEKEAAIRKVDEEAAKARQAINDATSNADVAAK 1236

Qy     467 IEEGGISAFSRSETP 481
Db    1237 QAQGTQAINNVPQTP 1251
```

RESULT 6
D84325
Htr17 transducer [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: D84325
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky,
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: D84325
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-536 <STO>
A;Cross-references: GB:AE004437; NID:g10581193; PIDN:AAG19968.1; GSPDB:GN00138
C;Genetics:
A;Gene: htr17
C;Superfamily: Halobacterium salinarum transducer protein htrI

Query Match 5.9%; Score 141; DB 2; Length 536;
Best Local Similarity 23.2%; Pred. No. 0.17;
Matches 119; Conservative 78; Mismatches 197; Indels 120; Gaps 23;

```
Qy      12 GVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAA 71
Db      63 GPVMDAFST----SLLAVLIGVVAISSTGTISIVED-LKTSNARAHR----RRKDAEQAK 113
```

```
Qy      72 LAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVL 131
           :||     | :|    :  |      |:   :| : |   :  :| : :    |||
Db     114 RRADEQRAQAQEAKAEAQDAREEAETLAAKLQERAAERDAQLSKKASEYSSVMEKAAS-- 171

Qy     132 QYMQPAINKGDWLATPLKPLTPLI-SGALSGAMDQVG----------TKMMDRAR----- 175
                 ||        ||     | : | || :||           :|: | |
Db     172 ---------GD--------LTVRANSSSDSDAMTEVGQAFNEMIAAFEEMVARTREFAGD 214

Qy     176 -----GDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPA 230
                |:|      :   : :| ||:  |    |:                  | || |
Db     215 VADNSRDIHASVDEIEAASEQVAESVQEISAGTDRE----------------HTRLADA 257

Qy     231 LASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSE 290
           |: |||:  |: :   ||:       :| |: |: | |:        :  ||:::
Db     258 TDELQSLSGAVE-EVAASASQVAD-------VSDQAADRGRQGS-------DDAAAAVTQ 302

Qy     291 ETDWLDAYKAIKS-ASYSGAALNAGKRMAGLPLDVATD--GLKAVRSLVSATSLTKNGLA 347
             ||     :|: | |:    |      :|:  |  |: ||   | |: : : |    ::|
Db     303 MTDIQSETEAVVSEANALAAEAQEIRRVVDLINDI-TDQTHLLALNASIEAARADQDGDG 361

Qy     348 LAGGYAGVSKLQKMATKNITD----------SATKAAVSQLSNLVGSVGVFAGWTTAGLA 397
             |   | | :| :  |   ||          |:|:: |      :  :|   :|  |   |
Db     362 FAVVAAEVKELAE-ETNEATDEIETLVDNLLSSTQSVVDDSEQMAATVS--SGTETVENA 418

Qy     398 -----------TDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAISST 446
                      ||  | :|   | |   : |:|     || || : : ::|:| |  ||
Db     419 LTALEEIGDHVTD--VNEAVQQIDDATDAQAAS-----ADTTVSMIEEIEDIS-EQNSSE 470

Qy     447 GASLRSTVNNLRHRSAPEADIEEGGISAFSRSET 480
           ||: :     : ::|  :|: :|      |::|
Db     471 AASVSAAA---QQQTASLSDVTDGVTRLVDRADT 501


RESULT 7
H89960
hypothetical protein SA1577 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: H89960
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: H89960
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2186 <KUR>
A;Cross-references: GB:BA000018; PID:g13701552; PIDN:BAB42845.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA1577

  Query Match             5.8%;  Score 140.5;  DB 2;  Length 2186;
  Best Local Similarity   19.3%;  Pred. No. 1.3;
  Matches  105;  Conservative   94;  Mismatches  203;  Indels  143;  Gaps   21;

Qy      20 TASDASL----ASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHN 75
           ||:: ::     :: |  | | : : :||||   |               | ::|
Db     930 TAANTNILNANTNADVEQVKTNAIQGIQAITPATK-----------VKTDAKNA----- 972

Qy      76 EQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQ 135
             ||   ||: | :::  :  ||: |  |  :  |:   ||          ||   | :
Db     973 --IDKSAETQHNTIFNNND--ATLEEQQAAQQLLDQAVATAKQNIN-----AADTNQEVA 1023

Qy     136 PAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVS 195
             | ::|      ::| | : : | :         : |:||  :  :: :|    :    :
Db    1024 QAKDQGTQNIVVIQPATQVKTDARNA--------VNDKAREAITNINATPGATREEKQEA 1075

Qy     196 VKRHSPALGRQVVDMGIAVQTFSALNVVRT-------VLAPALASRPSVQGA-------- 240
           : | :    | : |:|:   | : :| :|         : | :  : :  |
```

```
Db    1076 INRVNTLKNRALTDIGV-TSTTAMVNSIRDDAVNQIGAVQPHVTKKQTATGVLNDLATAK 1134

Qy     241 ---------------------VDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAF---- 275
                                 ||  ::||   : |    |    |    ||   |
Db    1135 KQEINQNTNATTEEKQVALNQVDQELATA---INNINQADTNAEVDQAQQLGTKAINAIQ 1191

Qy     276 --------VLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMA--------- 318
                   |   ::   | |:|     ||    |:|: :   ||  :: |
Db    1192 PNIVKKPAALAQINQHYNAKLAEINATPDATNDEKNAAIN--TLNQDRQQAIESIKQANT 1249

Qy     319 GLPLD----VATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMA--TKNITDSATK 372
             :|    || : : ||:  |      :: :      | |:|  :    | | ||   :
Db    1250 NAEVDQAATVAENNIDAVQVDVVKKQAARDKIT-----AEVAKRIEAVKQTPNATDEEKQ 1304

Qy     373 AAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKA-ESFIQDKVKSTASSTTSYV----ADQ 427
           |||:|:: |                | |: :  ::   |:| :| :   : :    |:
Db    1305 AAVNQINQL---------------KDQAINQINQNQTNDQVDTTTNQAVNAIDNVEAEV 1348

Qy     428 TVK------LAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPEADIEEGGISAFSRSETP 481
             :|      : | ||:   :  :    || || |      |   | :|  ::|  :::|
Db    1349 VIKPKAIADIEKAVKEKQQQIDN----SLDSTDNEKEVASQALAKEKEKALAAIDQAQTN 1404

Qy     482 FQLRR 486
           |: :
Db    1405 SQVNQ 1409


RESULT 8
T12796
probable transglycosylase - Bacillus subtilis phage SPBc2
C;Species: Bacillus subtilis phage SPBc2
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
C;Accession: T12796; A69911
R;Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A;Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 propha
A;Reference number: Z17583
A;Accession: T12796
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2285 <LAZ>
A;Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025510; PIDN:AAC13005.1
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: A69911
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2285 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14053.1; PID:e1183582;
A;Experimental source: strain 168
C;Genetics:
A;Gene: yomI

  Query Match             5.8%;  Score 140;  DB 2;  Length 2285;
  Best Local Similarity   20.5%;  Pred. No. 1.5;
  Matches  128;  Conservative   80;  Mismatches  224;  Indels  192;  Gaps   26;

Qy      15 MESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSV-IGSPDERDAALA 73
             :| ::  :  :: : : |  |:    : ||: |||     | :  :|  || : |    :
```

```
Db    251 IELYQRQAQVNVQNLNTRYGSSMGSSNRQAVQDYLN----AVNSLNVSTGSNNIRSQIQS 306

Qy     74 HNEQIDAL---VETRANRLYSEGETPATIAETF----------------------AKA 106
          | |    |    :| ||:  | |   | : :||                      ::|
Db    307 LNMQFRELASNAQTAANQASSFG---AELTQTFKSMSTYLISGSLFYGAISGLKEMVSQA 363

Qy    107 EKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQV 166
          : | | |           :    :|   | :|: || |:  :  :  :          |:
Db    364 IEIDTLMTNIRRVMNEPDYKYNELL---QESIDLGDTLSNKITDILQMTGDFGRMGFDES 420

Qy    167 GTKMMDRARGDLHYLS-TSPDKLHDAMAVSVKRHSPALGRQV--------VDMGIAVQTF 217
              : :    |  :|  :||   : :  ::   : |    :           ||   || |
Db    421 ELSTLTKTAQVLQNVSDLTPDDTVNTLTAAMLNFNIAANDSISIADKLNEVDNNYAVTTL 480

Qy    218 SALNVVRTVLAPALASRPSVQGAVDFGVSTA--------GGLVAN------AGFGDRMLS 263
             | :|   | : ||   |:      | :||         | :| |       |  |:   |
Db    481 DLANSIRK--AGSTASTFGVELNDLIGYTTAIASTTRESGNIVGNSLKTIFARIGNNQSS 538

Qy    264 VQSRDQLRGGAFVLGMKDKEPKAALSE---ETDWL-DAYKAIKSASYSG----------- 308
          ::: :|:      | : |     :||   : | | || |   |   :|
Db    539 IKALEQIGISVKTAGGEAKSASDLISEVAGKWDTLSDAQKQNTSIGVAGIYQLSRFNAMM 598

Qy    309 ----AALNAGK---------------------------------------------- 315
              | || |
Db    599 NNFSIAQNAAKTAANSTGSAWSEQQKYADSLQARVNKLQNNFTEFAIAASDAFISDGLIE 658

Qy    316 --RMAGLPLDVATDGLKAV---RSLVSATS-----LTKNGLALAGG-YAGVSKL--QKMA 362
            : ||  |: :|  :|:|     |::| |     |:||   ||      |   :  : :|
Db    659 FTQAAGSLLNASTGVIKSVGFLPPLLAAVSTATLLLSKNTRTLASSLILGTRAMGQETLA 718

Qy    363 TKNITDSATKAAVSQ----------LSNLVGSVGVFAGWTTAGLATDPA-VKKA----- 405
          |  :     |:|||:           :| |||      ||     | :  |  |||
Db    719 TAGLEAGMTRAAVASRVLKTALRGLLVSTLVGGAFAALGWALESLISSFAEAKKAKDDFE 778

Qy    406 ----------------ESFIQ-----DKVKSTASSTTSYVADQ---TVKLAKT----VK 436
                          :  ||      ||| : | |:    :    | :||:|    ||
Db    779 QSQQTNVEAITTNKDSTDKLIQQYKELQKVKESRSLTSDEEQEYLQVTQQLAQTFPALVK 838

Qy    437 --DMSGEAISSTGASLRSTVNNLR 458
            |  | ||  |   |   : | :
Db    839 GYDSQGNAILKTNKELEKAIENTK 862


RESULT 9
G64887
probable tail fiber protein GP37 - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: G64887; T09189
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G64887
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1122 <BLAT>
A;Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74454.1; PID:g1787636;
A;Experimental source: strain K-12, substrain MG1655
R;Aiba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai, H.
.; Motomura, K.; Nakade, S.; Nakamura, Y.; Nashimoto, H.; Nishio, Y.; Oshima, T.; Saito,
moto, Y.; Horiuchi, T.
DNA Res. 3, 363-377, 1996
A;Title: A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the
A;Reference number: Z16603; MUID:97251357; PMID:9097039
A;Accession: T09189
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 3-1122 <AIB>
A;Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74454.1; PID:g1787636;

  Query Match              5.7%;  Score 138;  DB 2;  Length 1122;
  Best Local Similarity   21.1%;  Pred. No. 0.76;
  Matches  106;  Conservative   65;  Mismatches  221;  Indels  110;  Gaps   17;

Qy      3 INQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVI 62
          : : |:     ||  :      || ||:| |  :| :
Db    112 VEEVARNASAVAQNTAAAKKSASDASTSAREAATHAA---------------------- 148

Qy     63 GSPDERDAALAHNEQIDALVETRANRLYSEG--ETPATIA-ETFAKAEKFDRLATTASSA 119
            |  |:| | :         :  :   | |   | || | :: | ||     | |:: |
Db    149 ---DAADSARAASTSAGQAASSAQSASSSAGTASTKATEASKSAAAAESSKSAAATSAGA 205

Qy    120 FENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLH 179
           : :   |:: ||          ||     |  | : | |   :|  : |:
Db    206 AKTSETNASASLQS----------AATSASTATTKASEAATSARDAAASK--EAAKSSET 253

Qy    180 YLSTSPD------------KLHDAMAVSVKRHSPALGRQV-VDMGIAVQTFSALNVVRT 225
            |:|              |      : :   | |:     |        |: : |
Db    254 NASSSASSAASSATAAGNSAKAAKTSETNARSSETAAGQSASAAAGSKTAAASSASAAST 313

Qy    226 VLAPALASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPK 285
            | ||  :  : :  |:|     ||     |  :|          |       |  |
Db    314 SAGQASASATAAGKSAESAASSASTATTKAGEATEQASAAARS-------ASAAKTSETN 366

Qy    286 AALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNG 345
          |  ||  :   :: |   ::| | || :|     |    | ||     | :|  |||: :
Db    367 AKASETS--AESSKTAAASSASSAASSASS--ASASKDEATRQASAAKS--SATTASTKA 420

Qy    346 LALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKA 405
            ||                |||  || |: :             :|     : | |:|
Db    421 TEAAG-----------------SATAAAQSKSTA-----------ESAATRAETAAKRA 451

Qy    406 ESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNLRHRSAPE- 464
          |           ||:|   :    |:|:      | |  :::|  :::|  :|   |   :
Db    452 EDIASAVALEDASTTKKGI----VQLSSATNSTS-ETLAATPKAVKSAYDNAEKRLQKDQ 506

Qy    465 --ADIEEGG-----ISAFSRSE 479
              ||| : |     |:| |:::
Db    507 NGADIPDKGCFLNNINAVSKTD 528


RESULT 10
B98241
hypothetical protein AGR_L_1764 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: B98241
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.,
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: B98241
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1731 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89452.1; PID:g15159316; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1764
A;Map position: linear chromosome

  Query Match              5.7%;  Score 138;  DB 2;  Length 1731;
  Best Local Similarity   21.4%;  Pred. No. 1.4;
  Matches  116;  Conservative   76;  Mismatches  197;  Indels  152;  Gaps   23;

Qy      1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60
          ||| :   :     :  | |   |:::||::  :|       :||       : | :  |
Db    746 MHIAELGTED----LAKFTTKDIAAISSSAIAGLSA------EAIASLTTAQIAALNTQS 795

Qy     61 VIGSPDERDAALAHNEQIDALVETRANRL--------------YSEGETPATIAETFA- 104
          :      : |||    |::||  |: | |                :|  :  |  ::  |
Db    796 INALSTAQIAALT-TAQVEALTSTQVNALTSKQIAALSTDDIATFSTKDIAAINSDAIAG 854

Qy    105 -KAEKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLAT-PLKPLTPLISGALSGA 162
```

```
Db    855 LSAETIASLATGQIAALNVRSIAALSTVQIVA--------LTTAQVEALTTVQVGALSST 906

Qy    163 MDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVV-------------- 208

Db    907 --------------QLAVLST------DDIATFSTRDMAALGSSAIAGLSKDTVASLTTA 946

Qy    209 DMG--------------IAVQTFSALNVVRTVLAPALASRP------SVQGAVDFGVSTA 248

Db    947 QIGALSMAGISGLSTGQIAALTGDQLNVLTNTQIAALTSKQVAAFDVSDITALSTGQIAA 1006

Qy    249 GGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEE------TDWLDAYKAIK 302

Db   1007 LSAAGAAGLTTDQIAALSTDQV--GAMTSGQ-----IAALSAKQIAALGTDDIATFSTGD 1059

Qy    303 SASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGY---------- 352

Db   1060 IAALSSNA------VAGLSRDTVASLTTAQIAALSSAGISGLGTGQIAGLTSEQVNVLTN 1113

Qy    353 AGVSKL--QKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDP----AVKKAE 406

Db   1114 AQISALTSKQVAALEVTDIASLSA-AQIA-AIGAAGV------AGLTTDQIAALSISQVE 1165

Qy    407 SFIQDKVKSTASSTTSYVADQTVKL----------------AKTVKDMSGEAISSTGA 448

Db   1166 ALTSAQIAALNSKQIAALSADDLAIFTTAEMAAIGSGAISGLPASTIASLTTAQIAALGA 1225

Qy    449 S 449

Db   1226 A 1226
```

RESULT 11
AB3045
ice nucleation protein homolog [imported] - Agrobacterium tumefaciens (strain C58, Dupon
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AB3045
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AB3045
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1731 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL44776.1; PID:g17742414; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: ina
A;Map position: linear chromosome

Query Match 5.7%; Score 138; DB 2; Length 1731;
Best Local Similarity 21.4%; Pred. No. 1.4;
Matches 116; Conservative 76; Mismatches 197; Indels 152; Gaps 23;

```
Qy      1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60

Db    746 MHIAELGTED----LAKFTTKDIAAISSSAIAGLSA------EAIASLTTAQIAALNTQS 795

Qy     61 VIGSPDERDAALAHNEQIDALVETRANRL---------------YSEGETPATIAETFA- 104

Db    796 INALSTAQIAALT-TAQVEALTSTQVNALTSKQIAALSTDDIATFSTKDIAAINSDAIAG 854

Qy    105 -KAEKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLAT-PLKPLTPLISGALSGA 162

Db    855 LSAETIASLATGQIAALNVRSIAALSTVQIVA--------LTTAQVEALTTVQVGALSST 906

Qy    163 MDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVV-------------- 208
```

```
Db    907 --------------QLAVLST------DDIATFSTRDMAALGSSAIAGLSKDTVASLTTA 946

Qy    209 DMG--------------IAVQTFSALNVVRTVLAPALASRP------SVQGAVDFGVSTA 248

Db    947 QIGALSMAGISGLSTGQIAALTGDQLNVLTNTQIAALTSKQVAAFDVSDITALSTGQIAA 1006

Qy    249 GGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEE------TDWLDAYKAIK 302

Db   1007 LSAAGAAGLTTDQIAALSTDQV--GAMTSGQ-----IAALSAKQIAALGTDDIATFSTGD 1059

Qy    303 SASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGY---------- 352

Db   1060 IAALSSNA------VAGLSRDTVASLTTAQIAALSSAGISGLGTGQIAGLTSEQVNVLTN 1113

Qy    353 AGVSKL--QKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDP----AVKKAE 406

Db   1114 AQISALTSKQVAALEVTDIASLSA-AQIA-AIGAAGV------AGLTTDQIAALSISQVE 1165

Qy    407 SFIQDKVKSTASSTTSYVADQTVKL----------------AKTVKDMSGEAISSTGA 448

Db   1166 ALTSAQIAALNSKQIAALSADDLAIFTTAEMAAIGSGAISGLPASTIASLTTAQIAALGA 1225

Qy    449 S 449

Db   1226 A 1226
```

RESULT 12
B90835
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: B90835
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: B90835
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-971 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB35073.1; PID:g13361114; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1650

Query Match 5.6%; Score 135; DB 2; Length 971;
Best Local Similarity 22.0%; Pred. No. 0.97;
Matches 116; Conservative 57; Mismatches 238; Indels 116; Gaps 20;

```
Qy      3 INQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVI 62

Db    110 VEEVARNASAVAQNTAAAKKSASDASTSAREAAT------------------------- 143

Qy     63 GSPDERDAALAHNEQIDALVETRANRLYSEG--ETPATIAETFAKAEKFDRLATTASSAF 120

Db    144 HATDAADSARAASTSAGQAASSAQSASSSAGTASTKATEASKSAAAAESSKSAAATSAGA 203

Qy    121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180

Db    204 AKTSETNAAVSQ--QSA-------ATSASTATTKASEAASSARDASASKEAAKS-SETSA 253

Qy    181 LSTSPDKLHDAMAV--SVK---------RHSPALGRQVVDMGIAVQTFSAL--NVVRTVL 227

Db    254 ASSASSAASSATAAGNSAKAAKTSETNAKSSETAAEQSASAAAGSKTAAALSASAASTSA 313

Qy    228 APALASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAA 287

Db    314 GQASASATAAGKSAESAASSASTATTKAGEATEQASA-------AASSASAAKTSETNAK 366

Qy    288 LSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLA 347
```

```
Db    367 ASETS--AESSKTAAASSASSAASSASS--ASASKDEATRQASAAKS--SATTASTKATE 420

Qy    348 LAGGYAGVSKLQ----------KMATKNITD----------SATKAAVSQLSNLVGSVGV 387

Db    421 AAGSATAAAQSKSTAESAATRAETAAKRAEDIASAVALEDASTTKKGIVQLSSATNS--- 477

Qy    388 FAGWTTAGLATDP-AVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAISST 446

Db    478 ----TSESLAATPKAVKAAYELANGK----------YTAQDATTAQKGIVQLS-NATNST 522

Qy    447 GASLRSTVNNLR--------HRSAPEADIEEGGI-----SAFSRSET 480

Db    523 SEMLAATPKSVKAAYDLANGKYTAQDATTAQKGIVQLSSATNSASET 569
```

RESULT 13
C85693
probable membrane protein of prophage CP-933X Z1918 [imported] - Escherichia coli (strai
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: C85693
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85693
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-973 <STO>
A;Cross-references: GB:AE005174; NID:g12514847; PIDN:AAG56007.1; GSPDB:GN00145; UWGP:Z19
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1918

Query Match 5.6%; Score 135; DB 2; Length 973;
Best Local Similarity 22.0%; Pred. No. 0.97;
Matches 116; Conservative 57; Mismatches 238; Indels 116; Gaps 20;

```
Qy      3 INQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVI 62

Db    112 VEEVARNASAVAQNTAAAKKSASDASTSAREAAT---------- 145

Qy     63 GSPDERDAALAHNEQIDALVETRANRLYSEG--ETPATIAETFAKAEKFDRLATTASSAF 120

Db    146 HATDAADSARAASTSAGQAASSAQSASSSAGTASTKATEASKSAAAAESSKSAAATSAGA 205

Qy    121 ENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHY 180

Db    206 AKTSETNAAVSQ--QSA-------ATSASTATTKASEAASSARDASASKEAAKS-SETSA 255

Qy    181 LSTSPDKLHDAMAV--SVK---------RHSPALGRQVVDMGIAVQTFSAL--NVVRTVL 227

Db    256 ASSASSAASSATAAGNSAKAAKTSETNAKSSETAAEQSASAAAGSKTAAALSASAASTSA 315

Qy    228 APALASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAA 287

Db    316 GQASASATAAGKSAESAASSASTATTKAGEATEQASA-------AASSASAAKTSETNAK 368

Qy    288 LSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLA 347

Db    369 ASETS--AESSKTAAASSASSAASSASS--ASASKDEATRQASAAKS--SATTASTKATE 422

Qy    348 LAGGYAGVSKLQ----------KMATKNITD----------SATKAAVSQLSNLVGSVGV 387

Db    423 AAGSATAAAQSKSTAESAATRAETAAKRAEDIASAVALEDASTTKKGIVQLSSATNS--- 479

Qy    388 FAGWTTAGLATDP-AVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAISST 446

Db    480 ----TSESLAATPKAVKAAYELANGK----------YTAQDATTAQKGIVQLS-NATNST 524

Qy    447 GASLRSTVNNLR--------HRSAPEADIEEGGI-----SAFSRSET 480

Db    525 SEMLAATPKSVKAAYDLANGKYTAQDATTAQKGIVQLSSATNSASET 571
```

RESULT 14
AC0304
probable hemolysin YPO2490 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C;Accession: AC0304
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0304
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2535 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC91295.1; PID:g15980484; GSPDB:GN00175
C;Genetics:
A;Gene: YPO2490

Query Match 5.6%; Score 134; DB 2; Length 2535;
Best Local Similarity 23.0%; Pred. No. 4.3;
Matches 84; Conservative 66; Mismatches 168; Indels 48; Gaps 16;

```
Qy    126 AAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSP 185

Db   1383 ATSEHIRHLGSEVNVGGALTANVDNLT-AVGANINAATLEVQAQNIS--------LSAAT 1433

Qy    186 DKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGV 245

Db   1434 DSLHVTGESSSKRHTSSV--NLYDETLLGSQLNATGDINLQAAQDITLRAS-------AV 1484

Qy    246 STAGGLVANAGFGDRMLSVQS--RDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKAIKS 303

Db   1485 QTDGALTLAAG-GDVLLTTQTEQHDEQRNHT---GLSKGIASSTLTRTEDSLSQTLAVGS 1540

Qy    304 ASYSGAALNAGKRMAGLPLDVATD---GLKAVRSLVSATS----------LTKNGLALA 349

Db   1541 MLSAGSIDVSGKNIAVMGSNVVADQDISLRAQENITVGTAQQSESESHLFEQKKSGLMST 1600

Qy    350 GGYAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFI 409

Db   1601 GG-IGVT-VGSSSTK-MTDSG--QSISSVGSTVGSV---LGNVSMTAGEDLRVQGAEVLA 1652

Qy    410 QDKVKSTASSTTSYVADQTVKLAKTVKD-MSGEAISSTGASLRSTVNNLRHRSAPEADIE 468

Db   1653 GKDINLTGKNVSILAAENQLTQSHTVEQKQSGLTLALSGA-VGSAVNTAVTTAKAASEES 1711

Qy    469 EGGISA 474

Db   1712 SGRLGA 1717
```

RESULT 15
C87587
hypothetical protein CC2730 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: C87587
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: C87587
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-654 <STO>

A;Cross-references: GB:AE005673; NID:g13424321; PIDN:AAK24695.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2730

Query Match 5.5%; Score 132.5; DB 2; Length 654;
Best Local Similarity 24.4%; Pred. No. 0.81;
Matches 116; Conservative 67; Mismatches 162; Indels 131; Gaps 28;

```
Qy      61 VIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAET----FAK--AEKFDRLAT 114
           |:||          |  : :|  ||   |     : ||::|     |||  |::   ::
Db       3 VVGS---------QNTALSSLSPTRIAAL-----SAATLSEAPASEFAKLTAQQVAAISI 48

Qy     115 TASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLT----PLISGALSGAMDQVGTKM 170
           || ||  :  |||  | |                 :: ||    | :| |:| :  |: |
Db      49 TAISALTSEQFAAFDVRQ--------------VRALTSAQLPRLSAAISFSPKQLAT-- 91

Qy     171 MDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGR-QVVDMGIAVQT-FSALNV------ 222
                  |    |   :     |:|:     |||  :   : :   | |:||::
Db      92 -------LSQAQTQALQHTQVAALSI----TALGEMRTTQLAVLSATAFAALSIDQVQSL 140

Qy     223 ----VRTVLAPALASRPSVQGAVDFGVSTAGGLVA--NAGFGDRMLSVQSRDQLRGGAFV 276
               ||:: || : : |:   | :| :   | | |   |       :|  | :|
Db     141 SLAQVRSLSAPQVRALPAADIA-EFSLEQVGALTALQLAVLSPTQVSALSSEQ------- 192

Qy     277 LGMKDKEPKAALSEET--------------DWLDAYKAIKSASYSGAALNA--GKRMAGL 320
           :|   :   :|||  |              |   |    : :| ||  | |    |:: |
Db     193 IGALSRTQFSALSAATVRALSPQTLALTTADHWRAATIAQMSSLSGDQLKALGETRLSSL 252

Qy     321 PLDVATDGLKAVRSLVSATSLTKNGLALAGG-----YAGVSKLQKMATKNITDSATKAAV 375
                 ||:| ||| : : | ||          ||:  :   |   :|| |  :|
Db     253 -------SASQVRAL-SATEVGRAGPALLDALTPTQLAGLGAIH-TAALTVTDIAGLSA- 302

Qy     376 SQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKL-AKT 434
           ||:: |       || :  ||:                :||   :  | |  :| |
Db     303 SQIAAL--DPAAVAGLSATGLSA---------------LSASQVAALSASQVARLRANQ 344

Qy     435 VKDMSGEAISS-TGASLRS-TVNNLRH-RSAPEADIEEGGISAFSRSETPFQLRRL 487
           :: ::||  |  | | |:: |   :|   ||  |:: :     |    |  |:| |
Db     345 LQGLNGEDFSEFTRAQLKAFTAAQIRELPSARLAELSDEDFGEF----TATQIRAL 396
```

Search completed: January 2, 2003, 15:21:21
Job time : 26.0175 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 2, 2003, 15:17:36 ; Search time 13.5139 Seconds
(without alignments)
1060.315 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINQSAQQPPGVAMESFRT.....EEGGISAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	146.5	6.1	2123	4	US-08-968-685A-10	Sequence 10, Appl
2	140	5.8	2285	4	US-09-308-375-2	Sequence 2, Appli
3	132	5.5	1026	2	US-08-614-377A-7	Sequence 7, Appli
4	132	5.5	1026	4	US-09-142-648B-7	Sequence 7, Appli
5	130	5.4	1026	1	US-08-194-290-7	Sequence 7, Appli
6	125.5	5.2	1222	2	US-08-682-517-15	Sequence 15, Appl
7	125.5	5.2	1252	2	US-08-682-517-9	Sequence 9, Appli
8	123	5.1	688	3	US-09-141-047-8	Sequence 8, Appli
9	122.5	5.1	1040	4	US-09-134-001C-5365	Sequence 5365, Ap
10	121.5	5.0	1068	4	US-09-085-199B-11	Sequence 11, Appl
11	117	4.9	682	1	US-08-441-139-2	Sequence 2, Appli
12	116	4.8	488	4	US-08-809-326A-1	Sequence 1, Appli
13	116	4.8	649	4	US-08-809-326A-15	Sequence 15, Appl
14	116	4.8	2887	4	US-08-462-467B-8	Sequence 8, Appli
15	115.5	4.8	401	2	US-08-591-079-6	Sequence 6, Appli
16	115	4.8	679	1	US-08-441-139-5	Sequence 5, Appli
17	114.5	4.8	748	4	US-09-066-046-4	Sequence 4, Appli
18	110.5	4.6	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
19	110	4.6	2089	1	US-08-418-893D-23	Sequence 23, Appl
20	110	4.6	2089	1	US-08-418-893D-24	Sequence 24, Appl
21	109.5	4.5	1618	4	US-08-462-467B-4	Sequence 4, Appli
22	109.5	4.5	2887	4	US-08-462-467B-2	Sequence 2, Appli
23	108.5	4.5	1002	4	US-09-268-347-24	Sequence 24, Appl
24	108	4.5	2314	4	US-09-268-347-49	Sequence 49, Appl
25	108	4.5	5087	4	US-09-144-085-1	Sequence 1, Appli
26	107.5	4.5	676	4	US-09-085-199B-7	Sequence 7, Appli
27	107.5	4.5	1140	4	US-09-220-081-2	Sequence 2, Appli
28	107.5	4.5	1140	4	US-09-677-575-2	Sequence 2, Appli
29	107.5	4.5	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
30	107	4.4	1004	4	US-09-268-347-30	Sequence 30, Appl
31	107	4.4	7257	3	US-09-335-409-5	Sequence 5, Appli
32	107	4.4	7257	4	US-09-568-102-5	Sequence 5, Appli
33	107	4.4	7257	4	US-09-567-969-5	Sequence 5, Appli
34	107	4.4	7257	4	US-09-568-480-5	Sequence 5, Appli
35	107	4.4	7257	4	US-09-568-486-5	Sequence 5, Appli
36	107	4.4	7257	4	US-09-568-472-5	Sequence 5, Appli
37	107	4.4	7257	4	US-09-567-899-5	Sequence 5, Appli
38	106	4.4	1833	4	US-08-621-944A-4	Sequence 4, Appli
39	106	4.4	1833	4	US-08-945-567D-4	Sequence 4, Appli
40	106	4.4	1992	4	US-08-621-944A-3	Sequence 3, Appli
41	106	4.4	1992	4	US-08-945-567D-3	Sequence 3, Appli
42	104.5	4.3	367	3	US-09-141-047-10	Sequence 10, Appl
43	104	4.3	547	2	US-08-467-822-35	Sequence 35, Appl
44	104	4.3	547	4	US-08-432-697-35	Sequence 35, Appl
45	104	4.3	547	4	US-08-466-248-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-08-968-685A-10
; Sequence 10, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSILA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2123 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-968-685A-10

Query Match 6.1%; Score 146.5; DB 4; Length 2123;
Best Local Similarity 22.2%; Pred. No. 0.00012;
Matches 111; Conservative 69; Mismatches 181; Indels 139; Gaps 23;

Qy 32 RSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDALVETR---ANR 88
Db 377 KTLTITGGAQTSALTD---------HNIGVVQNGDGLKVQLAETLTSLKMVTTENLTANE 427

```
Qy      89 LYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPL 148
Db     428 KVTVGKT---------------RL-TTDKIGFTND----MNGIDESKPYLDKDTGIHAGG 467

Qy     149 KPLTPLISGAL-----------------SGAMDQVGTKMMDRARGDLHYLSTSPDKLHDA 191
Db     468 QKITKLTAGVVDDDAATYGQLKKVNQTAESALQTFTVKKVDKNGND----------ANDS 517

Qy     192 MAVSVKRHSPALGRQVVDM------GIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGV 245
Db     518 KIITVGKNNKPDGTQVNTLKLKGENGVDVTT-----------------ETNGTVTFGL 558

Qy     246 STAGGL-VANAGFGDRMLSVQ---SRDQLRGGAFVLGMKD---KEPKAALSEET----DW 294
Db     559 NQNNGLTVGNSTLNNDGLSVKNTNSNKQIQVGADGITFTDISNSKPGAGIENTTRITRDG 618

Qy     295 LDAYKAIKSASYSGAALNAGKRMAGLPLDVATD---GLKAVRSLVSATSLTKNGLALAGG 351
Db     619 I-----------GFANNTGSLDANKPRLTPTGINAGGKELTNVQSAINPATNG------ 660

Qy     352 YAGVSKLQKMATKNITDSATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAV-KKAESFIQ 410
Db     661 -GQLDFMNRLSTANTEKSGSAATIKDLYNLSQVPLTFAG------DTGPNVTKKLGEILK 713

Qy     411 DK-VKSTASSTT----SYVADQ-----TVKLAKTVKDMSGEAISSTGASLRSTVNNLRHR 460
Db     714 VKGGKTTADDLTKNNIGVVADSTDNSLTVKLAKTLSDLDAVNTKTLTASDKVTVDSGNN- 772

Qy     461 SAPEADIEEGGISAFSRSET 480
Db     773 ---TAKLQNGDLT-FSKQNT 788
```

RESULT 2
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCT
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match 5.8%; Score 140; DB 4; Length 2285;
Best Local Similarity 20.5%; Pred. No. 0.00058;
Matches 128; Conservative 80; Mismatches 224; Indels 192; Gaps 26;

```
Qy      15 MESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSV-IGSPDERDAALA 73
Db     251 IELYQRQAQVNVQNLNTRYGSSMGSSNRQAVQDYLN----AVNSLNVSTGSNNIRSQIQS 306

Qy      74 HNEQIDAL---VETRANRLYSEGETPATIAETF----------------------AKA 106
Db     307 LNMQFRELASNAQTAANQASSFG---AELTQTFKSMSTYLISGSLFYGAISGLKEMVSQA 363

Qy     107 EKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPLTPLISGALSGAMDQV 166
Db     364 IEIDTLMTNIRRVMNEPDYKYNELL---QESIDLGDTLSNKITDILQMTGDFGRMGFDES 420

Qy     167 GTKMMDRARGDLHYLS-TSPDKLHDAMAVSVKRHSPALGRQV--------VDMGIAVQTF 217
Db     421 ELSTLTKTAQVLQNVSDLTPDDTVNTLTAAMLNFNIAANDSISIADKLNEVDNNYAVTTL 480

Qy     218 SALNVVRTVLAPALASRPSVQGAVDFGVSTA--------GGLVAN------AGFGDRMLS 263
Db     481 DLANSIRK--AGSTASTFGVELNDLIGYTTAIASTTRESGNIVGNSLKTIFARIGNNQSS 538

Qy     264 VQSRDQLRGGAFVLGMKDKEPKAALSE---ETDWL-DAYKAIKSASYSG---------- 308
Db     539 IKALEQIGISVKTAGGEAKSASDLISEVAGKWDTLSDAQKQNTSIGVAGIYQLSRFNAMM 598

Qy     309 ----AALNAGK---------------------------------------------- 315
Db     599 NNFSIAQNAAKTAANSTGSAWSEQQKYADSLQARVNKLQNNFTEFAIAASDAFISDGLIE 658

Qy     316 --RMAGLPLDVATDGLKAV---RSLVSATS-----LTKNGLALAGG-YAGVSKL--QKMA 362
Db     659 FTQAAGSLLNASTGVIKSVGFLPPLLAAVSTATLLLSKNTRTLASSLILGTRAMGQETLA 718

Qy     363 TKNITDSATKAAVSQ----------LSNLVGSVGVFAGWTTAGLATDPA-VKKA----- 405
Db     719 TAGLEAGMTRAAVASRVLKTALRGLLVSTLVGGAFAALGWALESLISSFAEAKKAKDDFE 778

Qy     406 ----------------ESFIQ-----DKVKSTASSTTSYVADQ---TVKLAKT----VK 436
Db     779 QSQQTNVEAITTNKDSTDKLIQQYKELQKVKESRSLTSDEEQEYLQVTQQLAQTFPALVK 838

Qy     437 --DMSGEAISSTGASLRSTVNNLR 458
Db     839 GYDSQGNAILKTNKELEKAIENTK 862
```

RESULT 3
US-08-614-377A-7
; Sequence 7, Application US/08614377A
; Patent No. 5976864
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 5976864ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
; TITLE OF INVENTION: HETEROLOGOUS
; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson PC
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/614,377A
; FILING DATE: 12-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; APPLICATION NUMBER: US 07/895,367
; FILING DATE: 09-JUNE-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsao, Y. Rocky
; REGISTRATION NUMBER: 34053
; REFERENCE/DOCKET NUMBER: 08106/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 7:

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1026 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-614-377A-7

  Query Match             5.5%;  Score 132;  DB 2;  Length 1026;
  Best Local Similarity   21.0%;  Pred. No. 0.00092;
  Matches  101;  Conservative   75;  Mismatches  166;  Indels  140;  Gaps   23;

Qy      22 SDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDAL 81
           |||:  ::::: |::|:   :|       : |     | |              :| |
Db      42 SDAAALTNTLKLVNSTTAVAIQT------YQFFTGVAPSAAG-------------LDFL 81

Qy      82 VETRAN------RLYSEGETPATIAETFAKAEKF----DRLATTASSAFENTPFAAA--- 128
           |::  |        ||:          ||:  :|      ||| | :    | ||||
Db      82 VDSTTNTNDLNDAYYSK----------FAQENRFINFSINLATGAGAG--ATAFAAAYTG 129

Qy     129 -SVLQYMQPAINK--GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSP 185
            |  | :  | :|  |: :||             : :|          : :: ||:
Db     130 VSYAQTVATAYDKIIGNAVAT------------AAGVDVAAAVAFLSRQANIDYLT--- 173

Qy     186 DKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVL---APALASRPSVQGAVD 242
                 |: ::|      :|: :       :    ||       | |:   :    |
Db     174 --------AFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAAMINDLSD 225

Qy     243 FGVST--AGGL-----VANAGFGDRMLSVQS-RDQLRG------------GAFVLGMKD 281
             :||  | |:        ::|     ||: :  | | |              ||  | : |
Db     226 GALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVAGEVAGAATLTVGD 285

Qy     282 KEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSL 341
                 | ::  :|: |            ||: |      ||  |    |:: : :: |  ::
Db     286 TLSGGAGTDVLNWVQA----------AAVTA------LPTGVTISGIETM-NVTSGAAI 327

Qy     342 TKNGLALAGGYAGVSKL----------------QKMATKNITDSATKAAVSQLSNL-VGS 384
           | |    : |  |:: |                | :       :|   ||   :|: | |
Db     328 TLN---TSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGRANVTVAS 384

Qy     385 VGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAIS 444
           || :| || |     |   |   :   | :::::||  :|  |   | ||   :| |::
Db     385 TGVTSGTTTVG-----ANSAASGTVSVSVANSSTTTTGAIA-VTGGTAVTVAQTAGNAVN 438

Qy     445 ST 446
           :|
Db     439 TT 440


RESULT 4
US-09-142-648B-7
; Sequence 7, Application US/09142648B
; Patent No. 6210948
; GENERAL INFORMATION:
;  APPLICANT:  Smit, John
;  APPLICANT:  Bingle, Wade H.
;  APPLICANT:  No. 6210948ellini, John F.
;  TITLE OF INVENTION:  EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
;  FILE REFERENCE:  08106/002002
;  CURRENT APPLICATION NUMBER: US/09/142,648B
;  CURRENT FILING DATE:  1999-03-30
;  PRIOR APPLICATION NUMBER:  PCT/CA97/00167
;  PRIOR FILING DATE:  1997-03-10
;  PRIOR APPLICATION NUMBER:  US 07/614,377
;  PRIOR FILING DATE:  1996-03-12
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 1026
;   TYPE: PRT
;   ORGANISM: Caulobacter crescentus
US-09-142-648B-7

  Query Match             5.5%;  Score 132;  DB 4;  Length 1026;
  Best Local Similarity   21.0%;  Pred. No. 0.00092;
  Matches  101;  Conservative   75;  Mismatches  166;  Indels  140;  Gaps   23;

Qy      22 SDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDAL 81
           |||:  ::::: |::|:   :|       : |     | |              :| |
Db      42 SDAAALTNTLKLVNSTTAVAIQT------YQFFTGVAPSAAG-------------LDFL 81

Qy      82 VETRAN------RLYSEGETPATIAETFAKAEKF----DRLATTASSAFENTPFAAA--- 128
           |::  |        ||:          ||:  :|      ||| | :    | ||||
Db      82 VDSTTNTNDLNDAYYSK----------FAQENRFINFSINLATGAGAG--ATAFAAAYTG 129

Qy     129 -SVLQYMQPAINK--GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSP 185
            |  | :  | :|  |: :||             : :|          : :: ||:
Db     130 VSYAQTVATAYDKIIGNAVAT------------AAGVDVAAAVAFLSRQANIDYLT--- 173

Qy     186 DKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVL---APALASRPSVQGAVD 242
                 |: ::|      :|: :       :    ||       | |:   :    |
Db     174 --------AFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAAMINDLSD 225

Qy     243 FGVST--AGGL-----VANAGFGDRMLSVQS-RDQLRG------------GAFVLGMKD 281
             :||  | |:        ::|     ||: :  | | |              ||  | : |
Db     226 GALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVAGEVAGAATLTVGD 285

Qy     282 KEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSL 341
                 | ::  :|: |            ||: |      ||  |    |:: : :: |  ::
Db     286 TLSGGAGTDVLNWVQA----------AAVTA------LPTGVTISGIETM-NVTSGAAI 327

Qy     342 TKNGLALAGGYAGVSKL----------------QKMATKNITDSATKAAVSQLSNL-VGS 384
           | |    : |  |:: |                | :       :|   ||   :|: | |
Db     328 TLN---TSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGRANVTVAS 384

Qy     385 VGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAIS 444
           || :| || |     |   |   :   | :::::||  :|  |   | ||   :| |::
Db     385 TGVTSGTTTVG-----ANSAASGTVSVSVANSSTTTTGAIA-VTGGTAVTVAQTAGNAVN 438

Qy     445 ST 446
           :|
Db     439 TT 440


RESULT 5
US-08-194-290-7
; Sequence 7, Application US/08194290
; Patent No. 5500353
;  GENERAL INFORMATION:
;    APPLICANT:  Smit, John
;    APPLICANT:  Bingle, Wade H
;    TITLE OF INVENTION:  Bacterial surface protein expression
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shlesinger, Arkwright & Garvey
;      STREET:  3000 South Eads Street
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/194,290
;      FILING DATE:  09-FEB-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Garvey, George A
;      REGISTRATION NUMBER:  17737
;      REFERENCE/DOCKET NUMBER:  5946-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-684-5600
;      TELEFAX:  703-836-5288
```

```
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1026 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-194-290-7

  Query Match             5.4%;  Score 130;  DB 1;  Length 1026;
  Best Local Similarity   21.0%;  Pred. No. 0.0014;
  Matches  101;  Conservative   75;  Mismatches  166;  Indels  140;  Gaps   23;

Qy      22 SDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDAL 81
           |||:  ::::: |::|:   :|       :  |      |  |              :| |
Db      42 SDAAALTNTLKLVNSTTAVAIQT------YQFFTGVAPSAAG-------------LDFL 81

Qy      82 VETRAN------RLYSEGETPATIAETFAKAEKF----DRLATTASSAFENTPFAAA--- 128
           |::  |        ||:          ||:  :|      |||  | :    | ||||
Db      82 VDSTTNTNDLNDAYYSK----------FAQENRFINFSINLATGAGAG--ATAFAAAYTG 129

Qy     129 -SVLQYMQPAINK--GDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSP 185
            |  | :  | :|  |: :||             :  :|          : ::  ||:
Db     130 VSYAQTVATAYDKIIGNAVAT------------AAGVDVAAAVAFLSRQANIDYLT--- 173

Qy     186 DKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFSALNVVRTVL---APALASRPSVQGAVD 242
                  |: ::|      :|: :       :    ||      | |:   :    |
Db     174 --------AFVRANTPFTAAADIDLAVKAALIGTILNAATVSGIGGYATATAAMINDLSD 225

Qy     243 FGVST--AGGL-----VANAGFGDRMLSVQS-RDQLRG------------GAFVLGMKD 281
             :||  | |:        ::|      ||: :  | | |              ||  | : |
Db     226 GALSTDNAAGVNLFTAYPSSGVSGSTLSLTTGTDTLTGTANNDTFVAGEVAGAATLTVGD 285

Qy     282 KEPKAALSEETDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSL 341
                | ::  :|: |            ||: |      ||  |   |:: : ::  |  ::
Db     286 TLSGGAGTDVLNWVQA----------AAVTA------LPTGVTISGIETM-NVTSGAAI 327

Qy     342 TKNGLALAGGYAGVSKL----------------QKMATKNITDSATKAAVSQLSNL-VGS 384
           | |    : |  |:: |                | :       :|   ||   :|: | |
Db     328 TLN---TSSGVTGLTALNTNTSGAAQTVTAGAGQNLTATTAAQAANNVAVDGGANVTVAS 384

Qy     385 VGVFAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKDMSGEAIS 444
           || :| || |     |   |   :   | :::::||  :|  |   | ||   :| |::
Db     385 TGVTSGTTTVG-----ANSAASGTVSVSVANSSTTTTGAIA-VTGGTAVTVAQTAGNAVN 438

Qy     445 ST 446
           :|
Db     439 TT 440


RESULT 6
US-08-682-517-15
; Sequence 15, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1222 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-682-517-15

  Query Match             5.2%;  Score 125.5;  DB 2;  Length 1222;
  Best Local Similarity   20.3%;  Pred. No. 0.0053;
  Matches  117;  Conservative   73;  Mismatches  219;  Indels  167;  Gaps   25;

Qy       3 INQSAQQPPGVAME-----------------------SFRTASDASLASSSVRSVSTT 37
           :||| ::  | :::                       |:    :|:|:|:       |
Db     675 VNQSGKKVTGTSIKKATYTIYNTGANDIKVDNQVISPNRSYTVTYEATLSST---GTVIT 731

Qy      38 SCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDALVET------------- 84
             ::|:  :  |          :  : | :| |    |       |
Db     732 PAKNLEVTSVDGKTTAVKVIATGIAVNTDGKDYAFTAKEATATFTATNEVPNSYTGVATQ 791

Qy      85 --------RANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQY--- 133
                   :| ::  |: |    |    |  |:       :  |    : |  : ||
Db     792 FNTADSGSNSNSIWFAGKNPVKYAGVSGKTYKY--FGANGNEVFGEAAWEAL-LTQYATE 848

Qy     134 ------------------MQPAINKGDWLATPLKPLTPL--ISGALS------GAMD--- 164
                            :  |:|       |: | ||    :|||:       | :|
Db     849 GQKVTISYNVDGDTVTFKVISAVNSSTEAIKPVAPTTPAAPTTGALTLTPAAGGLVDLTT 908

Qy     165 QVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVK-RHSPALGRQVVDMGIAVQTFSALNVV 223
             |  :  |  ||:  :|:     |   ||:|   : :|   :|: |      |:
Db     909 ATNTLGISLADADLNVSATTV----DTATVSLKDSANNSLSLTLVETGANTGVFAT---- 960

Qy     224 RTVLAPALASRPSVQGAVDFGVSTAGGLVAN-------AGFGDRMLSVQSRDQLRGG--- 273
            || |  |:|             |||  |           ||  : : :  :  :  |
Db     961 -TVQAGTLSS------------LTAGTLTVTYADAKNAAGVAENITASVTLKKTTGAITS 1007

Qy     274 -AFVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAAL---NAGKRMAGLPLDVATDGL 329
             |  |:    | || : |             |:  |  |   |||  ::  |       |
Db    1008 DTFTQGVL---PSAATAAEYTSKSIAADYTFATGEGFTLNIDNAGAQVINLA------GK 1058

Qy     330 KAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITD------SATKAAVSQLSNLV- 382
           |  : :  | : |  | |   |       |:  |: |          | ::|:|: | |
Db    1059 KGAQGVADAINATFAGTATVSG-------DKVVIKSATTGVGSEVEVTFSSVNQVLNAVV 1111

Qy     383 -GSVGVFAGW-TTAGLATDPAVKKAESFIQDKVKST----------------ASSTTSY 423
            |   | ||   |       |:   |  : | |: |                |::| :
Db    1112 NGKDQVVAGTAATKAFTITTALSVGEKVVIDGVEYTAVAFGTAPTANTFVVESAANTLAS 1171

Qy     424 VADQTVKLAKTVKDMSGE---AISSTGA--SLRSTV 454
           ||||   || |:  ::       |:|||  :| |||
Db    1172 VADQAANLAATIDTLNTADKFTASATGATITLTSTV 1207


RESULT 7
US-08-682-517-9
; Sequence 9, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1252 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-682-517-9

  Query Match             5.2%;  Score 125.5;  DB 2;  Length 1252;
  Best Local Similarity   20.3%;  Pred. No. 0.0055;
```

```
  Matches  117;  Conservative   73;  Mismatches  219;  Indels  167;  Gaps   25;

Qy       3 INQSAQQPPGVAME-------------------------SFRTASDASLASSSVRSVSTT 37
           :|||  ::   |  :::                       |:    :|:|:|:       |
Db     705 VNQSGKKVTGTSIKKATYTIYNTGANDIKVDNQVISPNRSYTVTYEATLSST---GTVIT 761

Qy      38 SCRDLQAITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDALVET------------ 84
             ::|:  :   |           :  : | :| |    |          |
Db     762 PAKNLEVTSVDGKTTAVKVIATGIAVNTDGKDYAFTAKEATATFTATNEVPNSYTGVATQ 821

Qy      85 --------RANRLYSEGETPATIAETFAKAEKFDRLATTASSAFENTPFAAASVLQY--- 133
                  :| ::  |: |   |     |  |:       :  |    : |  : ||
Db     822 FNTADSGSNSNSIWFAGKNPVKYAGVSGKTYKY--FGANGNEVFGEAAWEAL-LTQYATE 878

Qy     134 -----------------MQPAINKGDWLATPLKPLTPL--ISGALS------GAMD--- 164
                           :  |:|       |: | ||    :|||:       | :|
Db     879 GQKVTISYNVDGDTVTFKVISAVNSSTEAIKPVAPTTPAAPTTGALTLTPAAGGLVDLTT 938

Qy     165 QVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVK-RHSPALGRQVVDMGIAVQTFSALNVV 223
             |  :  |  ||:  :|:     |   ||:|   : :|    :|:  |     |:
Db     939 ATNTLGISLADADLNVSATTV----DTATVSLKDSANNSLSLTLVETGANTGVFAT---- 990

Qy     224 RTVLAPALASRPSVQGAVDFGVSTAGGLVAN-------AGFGDRMLSVQSRDQLRGG--- 273
             ||  |   |:|             |||  |        ||  : : :. :  :  |
Db     991 -TVQAGTLSS------------LTAGTLTVTYADAKNAAGVAENITASVTLKKTTGAITS 1037

Qy     274 -AFVLGMKDKEPKAALSEETDWLDAYKAIKSASYSGAAL---NAGKRMAGLPLDVATDGL 329
              |  |:     | ||  : |              |:  |  |   |||  ::  |       |
Db    1038 DTFTQGVL---PSAATAAEYTSKSIAADYTFATGEGFTLNIDNAGAQVINLA------GK 1088

Qy     330 KAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITD------SATKAAVSQLSNLV- 382
             |  : :  | :  |  |  |     |        |:  |: |        | ::|:|: | |
Db    1089 KGAQGVADAINATFAGTATVSG-------DKVVIKSATTGVGSEVEVTFSSVNQVLNAVV 1141

Qy     383 -GSVGVFAGW-TTAGLATDPAVKKAESFIQDKVKST----------------ASSTTSY 423
             |   | ||   |        |:   |  :  | |: |                |::| :
Db    1142 NGKDQVVAGTAATKAFTITTALSVGEKVVIDGVEYTAVAFGTAPTANTFVVESAANTLAS 1201

Qy     424 VADQTVKLAKTVKDMSGE---AISSTGA--SLRSTV 454
             ||||   || |:  ::       |:|||  :| |||
Db    1202 VADQAANLAATIDTLNTADKFTASATGATITLTSTV 1237


RESULT 8
US-09-141-047-8
; Sequence 8, Application US/09141047A
; Patent No. 6043085
; GENERAL INFORMATION:
;  APPLICANT: Walker, David H.
;  APPLICANT: Yu, Xue-Jie
;  TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
;  TITLE OF INVENTION: Protein Gene
;  FILE REFERENCE: D6143
;  CURRENT APPLICATION NUMBER: US/09/141,047A
;  CURRENT FILING DATE: 1998-08-27
;  NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 8
;   LENGTH: 688
;   TYPE: PRT
;   ORGANISM: Ehrlichia canis
;   FEATURE:
;   OTHER INFORMATION: Amino acid sequence of 120 kDa immunoreactive
;   OTHER INFORMATION: protein.
US-09-141-047-8

  Query Match             5.1%;  Score 123;  DB 3;  Length 688;
  Best Local Similarity   21.1%;  Pred. No. 0.0036;
  Matches  114;  Conservative   78;  Mismatches  205;  Indels  144;  Gaps   26;

Qy       8 QQPPGVAMESFRTASDASLASSSV---RSVSTTS---------CRDLQ-AITDYLKHHVF 54
           :  | |  |  : | | |:  ||      || ||          ||| |: | ::
Db     105 ESTPEVKAEDLQPAVDGSIEHSSSEVGEKVSKTSKEESTPEVKAEDLQPAVDDSVE---- 160

Qy      55 AAHRFSVIG----------SPDER--------DAALAHNEQIDALVETRANRLYSEGET 95
             | | :|           :|: :         | :: |:    : |  : ::   | |
Db     161 --HSSSEVGEKVSETSKEENTPEVKAEDLQPAVDGSIEHS---SSEVGEKVSKTSKEEST 215

Qy      96 PATIAETF----------AKAEKFDRLATTASSAFENTPFAAASVLQYMQPAINKGDWLA 145
           |   ||             : :|  :::: |:    ||||    |  ||
Db     216 PEVKAEDLQPAVDDSVEHSSSEVGEKVSETSKE--ENTPEVKAEDLQ------------ 260

Qy     146 TPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSPDKLHD---AMAVSVKRHSPA 202
                 | : |::  :  :|| |: : :: :    ||   |  |   |:  ||:  |
Db     261 -------PAVDGSVEHSSSEVGKKVSETSKEE----STPEVKAEDLQPAVDDSVEHSSSE 309

Qy     203 LGRQVVDMGIAVQTFSALNVVRTVLAPALAS---RPSVQGAVDFGVSTAGGLVANAGFGD 259
           :| :| :       |           | : :   :|:| |:|:   |  |  |:     :
Db     310 VGEKVSETSKEENT-----------PEVRAEDLQPAVDGSVEHSSSEVGEKVSETSKEE 357

Qy     260 RMLSVQSRD-------QLRGGAFVLGMKDKE-PKAALSEETDWLDAYKAI-----KSASY 306
             |:: |         :   :  :| |  |  |   : |    |   |:       |:|
Db     358 STPEVKAEDLQPAVDSSIEHSSSEVGKKVSETSKEESTPEVKAEDLQPAVDGSVEHSSSE 417

Qy     307 SGAALNAGKRMAGLP------LDVATDG-----LKAVRSLVSATSLTKNGLALAGGYAGV 355
             |  ::   :    |      |  | ||        |   || ||  :|   :
Db     418 VGEKVSETSKEENTPEVKAEDLQPAVDGSVEHSSSEVGEKVSETSKEENTPEVK-----A 472

Qy     356 SKLQKMATKNITDSATKAA--VSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKV 413
             ||      ::  |:::    ||: |    :  | |        | | :|: : | : :||
Db     473 EDLQPAVDGSVEHSSSEVGEKVSETSKEESTPEVKA--EDLQPAVDGSVEHSSSEVGEKV 530

Qy     414 KSTA--SSTTSYVA-DQTVKLAKTVKDMS---GEAISSTGASLRSTVNNLRHRSAPEADI 467
             |:   ||    | |   :  :|:  |   || :| |                :  | ||
Db     531 SETSKEESTPEVKAEDLQPAVDGSVEHSSSEVGEKVSETS----------KEESTPEVKA 580

Qy     468 E 468
           |
Db     581 E 581


RESULT 9
US-09-134-001C-5365
; Sequence 5365, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5365
;   LENGTH: 1040
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5365

  Query Match             5.1%;  Score 122.5;  DB 4;  Length 1040;
  Best Local Similarity   19.3%;  Pred. No. 0.008;
  Matches   95;  Conservative   71;  Mismatches  177;  Indels  149;  Gaps   19;

Qy      45 ITDYLKHHVFAAHRFSVIGSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFA 104
             |||: : ::  | | :    |||               :||: :                |
Db      19 ITDF-QRNIRKAQRMAKTEIPDE-------------IETQVD---------------A 47

Qy     105 KAEKFDRLATTASSAFEN-------------TPFAAASVLQYMQPAINKGDWLATPLKP 150
             || |   || :  :                 | | | : :|   :|   :     |
Db      48 NISKFKRALNTAKAMAQRWREHTVDIDGNANPVKRAIAVVREKLQQLRDKEVDIKGNNNP 107

Qy     151 LTPLISGA---LSGAMDQV-----GTKMMDRARGDLHYLSTSPDKL---HDAMAVSVKRH 199
             |   : ||   |:   |:       |: | ||:     |  | |:     ||:|  ::
```

```
Db    108 LKRSVLGAKAMLATLHDKTVHVNFDTRGMTRAQVLTRALGQSLDEYGDKMDALATKIRTF 167

Qy    200 SPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGAV---DFGVSTAGGLVANAG 256

Db    168 GTVFGQQI--KGMLIASFQGLIPIIAGLVPAIMAVANALGVVAGGALGVAGAFGIAASGA 225

Qy    257 FGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSE-ETDWLDAYKAIKSASYSGAALNAGK 315

Db    226 FAFGAMAVSAIKMLNDGTLQATAQTRRYQASLEQVKSTWEGIIKQ-NQAQIFNTLSNA-- 282

Qy    316 RMAGLPLDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNI------TDS 369

Db    283 ------LDTVNVALGRMKPFL----------------AGISKGMEQASQSVLKWAQNSQT 320

Qy    370 ATK-------AAVSQLSNLVGSVGVFAG----------WTTAGLATDPAVKKAESF 408

Db    321 ASKFFNMMNTTGVKTFNTLLSAAGRFGDGLINVFTQLGPLFLWTAKGL--DNLGKKFQNW 378

Qy    409 I-----QDKVKSTASST--------------------------TSYVADQTVKLAKTV 435

Db    379 ANSVAGQNAIKSFIEYTQTNLPKIGQIFGNVFMGIGNLMKAFAQNSSNIFDWLVKMTAKF 438

Qy    436 KDMSGEAISSTG 447

Db    439 REWSEQVGKSEG 450


RESULT 10
US-09-085-199B-11
; Sequence 11, Application US/09085199B
; Patent No. 6235879
;  GENERAL INFORMATION:
;    APPLICANT:  Hayden. Michael R.
;    APPLICANT:  Hackam, Abigail
;    APPLICANT:  Huq, A.H.M. Mahbubul
;    APPLICANT:  Chopra, Vikramjit Singh
;    APPLICANT:  Kalchman, Michael
;    TITLE OF INVENTION:  Apoptosis Modulators That Interact with the
;    TITLE OF INVENTION:  Huntington's Disease Gene
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Oppedahl & Larson
;      STREET:  PO Box 5270
;      CITY:  Frisco
;      STATE:  CO
;      COUNTRY:  USA
;      ZIP:  80443-5270
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 Kb storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS DOS 5.0
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/085,199B
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Larson, Marina T.
;      REGISTRATION NUMBER:  32038
;      REFERENCE/DOCKET NUMBER:  UBC.P-013US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (970) 668-2050
;      TELEFAX:  (970) 668-2052
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1068
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  no
;    ORIGINAL SOURCE:
;      ORGANISM:  mouse
;    FEATURE:
;      OTHER INFORMATION:  Huntington-interacting protein -mHIP1a
US-09-085-199B-11

  Query Match             5.0%;  Score 121.5;  DB 4;  Length 1068;
  Best Local Similarity   22.7%;  Pred. No. 0.01;
  Matches  119;  Conservative   74;  Mismatches  178;  Indels  153;  Gaps   29;

Qy      1 MHINQSAQQPPGVAMESFRTASDASLASSSVRSVSTTSCRDLQAITDYLKHHVFAAHRFS 60

Db    505 MKMEEQSDQLEKLKRELAARAGELARAQEAL-SRTEQSGSELSSRLDTLN-----AEKEA 558

Qy     61 VIGSPDERDAALAH-----NEQIDALVETRANRLYSEGETPATIAETFAKAEKF------ 109

Db    559 LSGVVRQREAELLAAQSLVREKEEALSQEQQRSSQEKGELRGQLAEKESQEQGLRQKLLD 618

Qy    110 DRLATTASSAFENTPFAAASVLQYMQPAINKGDWLATPLKPL------TP--LISGALSG 161

Db    619 EQLAVLRS--------AAAEAEAILQDAVSKLD------DPLHLRCTSSPDYLVSRA-QA 663

Qy    162 AMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTFS--- 218

Db    664 ALDSVS----GLEQGHTQYLASSEDA--SALVAALTRFSHLAADTIVN-GAATSHLAPTD 716

Qy    219 ---------------ALNVV-----RTVLAPALAS--RPSVQGAVDFGVSTAGGLVANAG 256

Db    717 PADRLMDTCRECGARALELVGQLQDQTVLRRAQPSLMRAPLQGILQLGQDL--------- 767

Qy    257 FGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDAYKAIKSA------SYSGAA 310

Db    768 ---KPKSLDVRQEELG-----AMVDKE-MAATSAAIE--DAVRRIEDMMSQARHESSGVK 816

Qy    311 LNAGKRMAGLPLDVATDGLKAVRSLV-SATSLTKNGLALAGGYAGVSKLQKMATKNI--T 367

Db    817 LEVNERI----LNSCTDLMKAIRLLVMTSTSLQKE---IVESGRGAATQQEFYAKNSRWT 869

Qy    368 D---SATKAAVSQLSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQDKVK---------S 415

Db    870 EGLISASKA----------------VGWGATQL-----VESADKVVLHMGKYEELIVCSHE 909

Qy    416 TASSTTSYVADQTVK-------------LAKTVKDMSGEAISST 446

Db    910 IAASTAQLVAASKVKANKNSPHLSRLQECSRTVNERAANVVAST 953


RESULT 11
US-08-441-139-2
; Sequence 2, Application US/08441139
; Patent No. 5773245
;  GENERAL INFORMATION:
;    APPLICANT:  Wittrup, Dr. Karl D.
;    APPLICANT:  Robinson, Anne S.
;    TITLE OF INVENTION:  METHODS FOR INCREASING SECRETION OF
;    TITLE OF INVENTION:  RECOMBINANTLY EXPRESSED PROTEINS
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/441,139
;      FILING DATE:  15-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/089,997
;      FILING DATE:  06-JUL-1993
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DiGiglio, Frank S.
;      REGISTRATION NUMBER:  31,346
;      REFERENCE/DOCKET NUMBER:  8646
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  516-742-4343
;      TELEFAX:  516-742-4366
;      TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  682 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-441-139-2

  Query Match             4.9%;  Score 117;  DB 1;  Length 682;
  Best Local Similarity   21.2%;  Pred. No. 0.014;
  Matches  114;  Conservative   64;  Mismatches  165;  Indels  194;  Gaps   28;

Qy      71 ALAHNEQIDALVETRANRLYSEGETPATIAET-----FAKAEKFDRLATTASSAFE---- 121
           |:  | : : |   : ||:      ||: :| |        | |    |   :: |:
Db      65 AVMKNGKTEILANEQGNRI-----TPSYVAFTDDERLIGDAAKNQVAANPQNTIFDIKRL 119

Qy     122 ---------------NTPFAAASVLQYMQPAIN---KGDWLATPLKPLTP-LISGALSGA 162
                         : ||    | :  :||:    ||:      |  ||  ||| : |
Db     120 IGLKYNDRSVQKDIKHLPFNV--VNKDGKPAVEVSVKGE-----KKVFTPEEISGMILGK 172

Qy     163 MDQV-----GTKMMDRARGDLHYLSTSPDKLHDAMAVSVKRHSPALGRQVVDMGIAVQTF 217
           | |:     |||:         | : | |   :||   : |           | |    |
Db     173 MKQIAEDYLGTKV-------THAVVTVPAYFNDAQRQATK-----------DAG----TI 210

Qy     218 SALNVVRTVLAPALASRPSVQGAVDFGVSTAGG----LVANAG---FGDRMLSVQS---R 267
           : |||:| |   |  |        |: :|:  :      :| : |   |   :||:::
Db     211 AGLNVLRIVNEPTAA-------AIAYGLDKSDKEHQIIVYDLGGGTFDVSLLSIENGVFE 263

Qy     268 DQLRGGAFVLGMKD---------------------------------KEPKAALSEE- 291
           |   |   || :|                                 :: | ||| :
Db     264 VQATSGDTHLGGEDFDYKIVRQLIKAFKKKHGIDVSDNNKALAKLKREAEKAKRALSSQM 323

Qy     292 ---------TDWLDAYKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLT 342
                   | :|  : :  | :                | ||:    ||  |  ::  : |
Db     324 STRIEIDSFVDGIDLSETLTRAKFE-----------ELNLDLFKKTLKPVEKVLQDSGLE 372

Qy     343 K---NGLALAGGYAGVSKLQ---------KMATKNIT--DSATKAAVSQLSNLVGSVGV- 387
           |   : : | ||   : |:|          | |:| |   ::    |  |   | |  ||
Db     373 KKDVDDIVLVGGSTRIPKVQQLLESYFDGKKASKGINPDEAVAYGAAVQAGVLSGEEGVE 432

Qy     388 ----------FAGWTTAGLATDPAVKKAESFIQDKVKSTASSTTSYVADQTVKLAKTVKD 437
                      | | |   | :|:  : |  | ||   ||   | :|   : |  :
Db     433 DIVLLDVNALTLGIETTGGVMTPLIKR-NTAIPTK-KSQIFSTA--VDNQPTVMIKVYE- 487

Qy     438 MSGEAISSTGASLRSTVNNL-----------RHRSAPEADI----EEGGISAFSRSE 479
                 |     |  |||              | |: ::    :  ||   | ::
Db     488 ---------GERAMSKDNNLLGKFELTGIPPAPRGVPQIEVTFALDANGILKVSATD 535


RESULT 12
US-08-809-326A-1
; Sequence 1, Application US/08809326A
; Patent No. 6165478
;  GENERAL INFORMATION:
;    APPLICANT:  Izutsu, Hiroshi
;    APPLICANT:  Obara, Kazuhiko
;    APPLICANT:  Matsumoto, Akira
;    TITLE OF INVENTION:  CHLAMYDIA PNEUMONIAE ANTIGENIC POLYPEPTIDES,
;    TITLE OF INVENTION:  FUSED PROTEINS CONTAINING THE POLYPEPTIDES, DNAS CODING
;    TITLE OF INVENTION:  THEREFOR, RECOMBINANT VECTORS CARRYING THE DNAS, TRANSFORMANTS
;    TITLE OF INVENTION:  CONTAINING THE RECOMBINANT VECTORS, METHOD FOR PRODUCTION OF
;    TITLE OF INVENTION:  ANTIBODY, METHOD AND REAGENTS FOR DETECTION AND/OR MEASUREMENT
;    TITLE OF INVENTION:  OF ANTIBODY, METHOD AND AGENTS FOR DIAGNOSIS OF CHLAMYDIA
;    TITLE OF INVENTION:  PNEUMONIAE INFECTIONS, PROBES AND PRIMERS FOR DETECTION AND/OR
;    TITLE OF INVENTION:  MEASUREMENT OF CHLAMYDIA PNEUMONIAE GENE, AND METHOD AND
;    TITLE OF INVENTION:  REAGENTS FOR DETECTION AND/OR MEASUREMENT OF CHLAMYDIA
;    TITLE OF INVENTION:  PNEUMONIAE GENE
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/809,326A
;      FILING DATE:  19-MAR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 224711/94
;      FILING DATE:  20-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 106006/95
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 106008/95
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 106009/95
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 106010/95
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 106011/95
;      FILING DATE:  28-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Miller, Charles E.
;      REGISTRATION NUMBER:  24,576
;      REFERENCE/DOCKET NUMBER:  7426-043-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  488 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-809-326A-1

  Query Match             4.8%;  Score 116;  DB 4;  Length 488;
  Best Local Similarity   21.6%;  Pred. No. 0.0098;
  Matches   95;  Conservative   59;  Mismatches  157;  Indels  128;  Gaps   20;

Qy      63 GSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFEN 122
           | | :   :    :||    :||  :  :| |: ||||    | :      |::::  |
Db      26 GVPQQDKLSGNETKQIQ---QTRQGK-NTEMESDATIAGASGKDK------TSSTTKTET 75

Qy     123 TP---FAAASVLQYMQPA-INKGDWLATPLKPLTPLISGALSGAMDQVGTKM-------- 170
           |    ||       | |  : |             :||| :       ||:
Db      76 APQQGVAAGKESSESQKAGADTG------------VSGAAATTASNTATKIAMQTSIEE 122

Qy     171 ----MDRARGDLHYLSTSPDKLHDAMAV---------SVKRHSPALGR-------QVVDM 210
                |:     |  || :  |  :|: |          | |  :| | :       :|:::
Db     123 ASKSMESTLESLQSLSAAQMKEVEAVVVAALSGKSSGSAKLETPELPKPGVTPRSEVIEI 182

Qy     211 GIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSV-QSRDQ 269
           |:|:    |:  :      ||::  | |   |             |   : : : : |::
Db     183 GLALA--KAIQTLGEATKSALSNYASTQAQAD--------QTNKLGLEKQAIKIDKEREE 232
```

```
Qy     270 LRGGAFVLGMKDKEPKAALSEETDWLDAYK--------AIKSASYSGAALNAGKRMAGLP 321
Db     233 YQ------EMKAAEQKSKDLEGT--MDTVNTVMIAVSVAITVISIVAAIFTCGAGLAGLA 284

Qy     322 LDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNL 381
Db     285 AGAAV------------------GAAAAGGAAGAAAATTVATQ-ITVQAVVQAVKQ---- 321

Qy     382 VGSVGVFAGWTTAGLATDPAVKKA-----ESFIQDKVKSTASST-----------TSYVA 425
Db     322 -------AVITAVRQAITAAIKAAVKSGIKAFIKTLVKAIAKAISKGISKVFAKGTQMIA 374

Qy     426 DQTVKLAKTVKDMSGEAIS 444
Db     375 KNFPKLSKVISSLTSKWVT 393


RESULT 13
US-08-809-326A-15
; Sequence 15, Application US/08809326A
; Patent No. 6165478
;  GENERAL INFORMATION:
;    APPLICANT:  Izutsu, Hiroshi
;    APPLICANT:  Obara, Kazuhiko
;    APPLICANT:  Matsumoto, Akira
;    TITLE OF INVENTION:  CHLAMYDIA PNEUMONIAE ANTIGENIC POLYPEPTIDES,
;    TITLE OF INVENTION:  FUSED PROTEINS CONTAINING THE POLYPEPTIDES, DNAS CODING
;    TITLE OF INVENTION:  THEREFOR, RECOMBINANT VECTORS CARRYING THE DNAS, TRANSFORMANTS
;    TITLE OF INVENTION:  CONTAINING THE RECOMBINANT VECTORS, METHOD FOR PRODUCTION OF
;    TITLE OF INVENTION:  ANTIBODY, METHOD AND REAGENTS FOR DETECTION AND/OR MEASUREMENT
;    TITLE OF INVENTION:  OF ANTIBODY, METHOD AND AGENTS FOR DIAGNOSIS OF CHLAMYDIA
;    TITLE OF INVENTION:  PNEUMONIAE INFECTIONS, PROBES AND PRIMERS FOR DETECTION AND/OR
;    TITLE OF INVENTION:  MEASUREMENT OF CHLAMYDIA PNEUMONIAE GENE, AND METHOD AND
;    TITLE OF INVENTION:  REAGENTS FOR DETECTION AND/OR MEASUREMENT OF CHLAMYDIA
;    TITLE OF INVENTION:  PNEUMONIAE GENE
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/809,326A
;      FILING DATE:  19-MAR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 224711/94
;      FILING DATE:  20-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 106006/95
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 106008/95
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 106009/95
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 106010/95
;      FILING DATE:  28-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 106011/95
;      FILING DATE:  28-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Miller, Charles E.
;      REGISTRATION NUMBER:  24,576
;      REFERENCE/DOCKET NUMBER:  7426-043-999
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  649 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-809-326A-15

  Query Match             4.8%;  Score 116;  DB 4;  Length 649;
  Best Local Similarity   21.6%;  Pred. No. 0.016;
  Matches   95;  Conservative   59;  Mismatches  157;  Indels  128;  Gaps   20;

Qy      63 GSPDERDAALAHNEQIDALVETRANRLYSEGETPATIAETFAKAEKFDRLATTASSAFEN 122
Db     187 GVPQQDKLSGNETKQIQ---QTRQGK-NTEMESDATIAGASGKDK------TSSTTKTET 236

Qy     123 TP---FAAASVLQYMQPA-INKGDWLATPLKPLTPLISGALSGAMDQVGTKM-------- 170
Db     237 APQQGVAAGKESSESQKAGADTG------------VSGAAATTASNTATKIAMQTSIEE 283

Qy     171 ----MDRARGDLHYLSTSPDKLHDAMAV---------SVKRHSPALGR-------QVVDM 210
Db     284 ASKSMESTLESLQSLSAAQMKEVEAVVVAALSGKSSGSAKLETPELPKPGVTPRSEVIEI 343

Qy     211 GIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVANAGFGDRMLSV-QSRDQ 269
Db     344 GLALA--KAIQTLGEATKSALSNYASTQAQAD--------QTNKLGLEKQAIKIDKEREE 393

Qy     270 LRGGAFVLGMKDKEPKAALSEETDWLDAYK--------AIKSASYSGAALNAGKRMAGLP 321
Db     394 YQ------EMKAAEQKSKDLEGT--MDTVNTVMIAVSVAITVISIVAAIFTCGAGLAGLA 445

Qy     322 LDVATDGLKAVRSLVSATSLTKNGLALAGGYAGVSKLQKMATKNITDSATKAAVSQLSNL 381
Db     446 AGAAV------------------GAAAAGGAAGAAAATTVATQ-ITVQAVVQAVKQ---- 482

Qy     382 VGSVGVFAGWTTAGLATDPAVKKA-----ESFIQDKVKSTASST-----------TSYVA 425
Db     483 -------AVITAVRQAITAAIKAAVKSGIKAFIKTLVKAIAKAISKGISKVFAKGTQMIA 535

Qy     426 DQTVKLAKTVKDMSGEAIS 444
Db     536 KNFPKLSKVISSLTSKWVT 554


RESULT 14
US-08-462-467B-8
; Sequence 8, Application US/08462467B
; Patent No. 6210899
;  GENERAL INFORMATION:
;    APPLICANT:  Rosenbaum, Jan S
;    TITLE OF INVENTION:  The Use of a BMP Protein Receptor
;    TITLE OF INVENTION:  Complex for Screening Bone Metabolism Actives and Cells
;    TITLE OF INVENTION:  Co-Transfected With a Type II BMP Receptor and a Type I
;    TITLE OF INVENTION:  BMP Receptor
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Procter & Gamble Company
;      STREET:  11810 East Miami River Road
;      CITY:  Ross
;      STATE:  OH
;      COUNTRY:  USA
;      ZIP:  45061
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/462,467B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hersko, Bart S.
;      REGISTRATION NUMBER:  32,572
;      REFERENCE/DOCKET NUMBER:  5474R
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (513) 627-0633
;      TELEFAX:  (513) 627-0260
;  INFORMATION FOR SEQ ID NO:  8:.
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2887 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
US-08-462-467B-8

  Query Match             4.8%;  Score 116;  DB 4;  Length 2887;
  Best Local Similarity   21.0%;  Pred. No. 0.19;
  Matches  117;  Conservative   76;  Mismatches  217;  Indels  146;  Gaps   22;

Qy      24 ASLASSSVRSVS-TTSCRDLQAITDYLKHHVFAA---HRFSVIGSPDERDAALAHNEQID 79
           || ||::: :::    | | | |:  |  | :  |    :|     :|     ||     :
Db    1027 ASPASNALAALAILESERGLVALGLY-THRILEARGTYRMETALAP-RGLVALLEGLGLY 1084

Qy      80 ALVETRANRLYSEGETP--ATIAETFAKAEKFDRLATTASSAFENTPFAAA---SVLQYM 134
             ||    |: | : | :|  : ::|     | :  |    :|  | |  | |     |: :
Db    1085 ALAVALASNLEARGASPCYSGLSERALALELYSGLNVALASPMETTYRALALEGLYLEIL 1144

Qy     135 QPAINKGDWLATPLKPLTPLISGALSGAMDQVGTKMMDRARGDLHYLSTSPDKLHDAMAV 194
           :   :  :| | :                  |: |||       || : |:   :
Db    1145 ETYRTRPGLVALPHE-------------------METARGCYSTHRASPLEPHEPRGL 1183

Qy     195 SVKRHSPALGRQVVDMGIAVQTFSALNVVRTVLAPALASRPSVQGAVDFGVSTAGGLVA- 253
             || |       |  |:  || : | |||              |::|  ||||
Db    1184 YGLSERVALPR-------ASPTYRGLN-METALAP-----------HEGLNTHRGLVAL 1223

Qy     254 ----------------------NAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPK----- 285
                                 |    :| |   :: |||    |:  : |:
Db    1224 GLYASNHISPRTHRPHEGLASPMETGLNVALLEVALSERARGGLLYSGLNARGPRLYSPH 1283

Qy     286 -------------------------AALSEETDWLDAYKAIKSASYSGAA-------- 310
                                     ||:  :: |: |  :       | |
Db    1284 EPRGLALATRPLYSGLASNSERLEALAVALARGSERLELYSGLTHRILEGLASPCYSTRP 1343

Qy     311 ----LNAGKRMAGLPLDVATDGLKAVRSLVSATSLTK-NGLALAGG----YAGVSKLQKM 361
               |||   :||| |  |  ||:  |:|      :   || || |     ||: ::: |
Db    1344 ASPGLNASPALAGLAL--AARGLETHRALAGLNCYSALAGLGLARGMETALAGLLEMETM 1401

Qy     362 ATKNITDSATKAAVSQ-LSNLVGSVGVFAG----WTTAGLATDPAVKKAESFIQDKVKST 416
           |  | :  |:  :::  |||  |  |            ||:|  :: :|   :     |
Db    1402 ET--ILE--TRPGLARGASNLYSSERVALSERPRTHRVALASNPRMETSERTHRALAMET 1457

Qy     417 ASSTTSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNNL--------RHRSAPEADIE 468
             : :: :|      |          :  | : | :|    : |        | |::|     |
Db    1458 GLNASNGLARGASNLESERHISASNARGARGVALPRLYSILEGLYPRTYRPRASPTYRSE 1517

Qy     469 EGGISAFSRSETPFQL 484
             |   |||  :::
Db    1518 RSERS--ERSERTYRI 1531


RESULT 15
US-08-591-079-6
; Sequence 6, Application US/08591079
; Patent No. 5972899
;  GENERAL INFORMATION:
;    APPLICANT:  Zychlinksky, Arturo
;    APPLICANT:  Chen, Yajing
;    TITLE OF INVENTION:  Apoptosis Induced by Shigella IpaB
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Avenue, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1812
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/591,079
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Livnat, Shmuel
;      REGISTRATION NUMBER:  33,949
;      REFERENCE/DOCKET NUMBER:  15661-20017.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0764
;      TELEX:  90-4030 MRSNFOERSWSH
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 401 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-591-079-6

  Query Match             4.8%;  Score 115.5;  DB 2;  Length 401;
  Best Local Similarity   21.2%;  Pred. No. 0.0079;
  Matches   88;  Conservative   51;  Mismatches  120;  Indels  157;  Gaps   21;

Qy     122 NTPFAAASVLQYMQ---PAINKGDWLATPLK---------------PLTPLISGALSGAM 163
           :||    |:| |::   ||  :   :|  ||                ||  : |    |
Db      10 STP-VTGSLLPYVETPAPAPLQTQQVAGELKDKNGGVSSQGVQLPAPLAVVASQVTEGQQ 68

Qy     164 DQVGTKMMDR-ARGDL------HYLS-------TSPD---------------------- 186
             :| ||:::   ||         :|:|         |||
Db      69 QEV-TKLLESVTRGAAGSQLISNYVSVLTKFTLASPDTFEIELGKLVSNLEEVRKDIKIA 127

Qy     187 ---KLHDAMAVSVKRHSPAL------GRQVVDMGIAVQTFSALNVVRTVLAPALASRPSV 237
              :||:      :: :   :        :||   ||| : |   |: : :|:   |:     |
Db     128 DIQRLHEQNMKKIEENQEKIKETEENAKQVKKSGIASKIFGWLSAIASVIVGAIMVASGV 187

Qy     238 QGAVDFGVSTAGGLVANAGFGDRMLSVQSRDQLRGGAFVLGMKDKEPKAALSEETDWLDA 297
             |||       || ::  :|                  |:||                |
Db     188 -GAV------AGAMMVASG------------------VIGM-----------------A 204

Qy     298 YKAIKSASYSGAALNAGKRMAGLPLDVATDGLKAVRSLVSATSLTKNGLAL-----AGGY 352
             |:| |:  |        :: | |:  | :     |    | :| :|  | ||       |
Db     205 NMAVKQAAEDGLISQEAMKILG-PILTAIE----VALTVVSTVMTFGGSALKCLANIGAK 259

Qy     353 AGVSKLQKMATKNITDSATKAAVSQ-LSNLVGSVGVFAGWTTAGLATDPAVKKAESFIQD 411
             | :    :| |     ||  | :|  :|| |||      | : |||     |::
Db     260 LG-ANTASLAAKGAEFSAKVAQISTGISNTVGSAVTKLGGSFAGLTMSHAIR-------- 310

Qy     412 KVKSTASSTTSYVADQTVKLAKTVKDMSGEAISSTGASLRSTVNN-----LRHRSA 462
             | |  |                    :     |: :  |:||     |:| :|
Db     311 ----TGSQAT-----------------QVAVGVGSGITQTINNKKQADLQHNNA 343


Search completed: January  2, 2003, 15:21:49
Job time : 21.5139 secs
```